AUTOMATED ANALYSIS OF X-RAY IMAGES FOR CARGO SECURITY

THOMAS W. ROGERS

A dissertation submitted in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY
of
UNIVERSITY COLLEGE LONDON

Department of Security and Crime Science

University College London

I·VI·MMXVII

DECLARATION

I, Thomas William Rogers, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

London, 1st June 2017

THOMAS W. ROGERS

It is said that your life flashes before your eyes just before you die.

That is true, it's called Life.

— Terry Pratchett

Dedicated to the loving memory of

THOMAS W. ROGERS

1960 – 2014

ABSTRACT

Customs and border officers are overwhelmed by the hundreds of millions of cargo containers that constitute the backbone of the global supply chain, any one of which could contain a security- or customs-related threat. Searching for these threats is akin to searching for needles in an ever-growing field of haystacks.

This thesis considers novel automated image analysis methods to automate or assist elements of cargo inspection. The four main contributions of this thesis are as follows.

Methods are proposed for the measurement and correction of detector wobble in large-scale transmission radiography using Beam Position Detectors (BPDs). Wobble is estimated from BPD measurements using a Random Regression Forest (RRF) model, Bayesian fused with a prior estimate from an Auto-Regression (AR). Next, a series of image corrections are derived, and it is shown that 87% of image error due to wobble can be corrected. This is the first proposed method for correction of wobble in large-scale transmission radiography.

A Threat Image Projection (TIP) framework is proposed, for training, probing and evaluating Automated Threat Detection (ATD) algorithms. The TIP method is validated experimentally, and a method is proposed to test whether algorithms can learn to exploit TIP artefacts.

A system for Empty Container Verification (ECV) is proposed. The system, trained using TIP, is based on Random Forest (RF) classification of image patches according to fixed geometric features and container location. The method outperforms previous reported results, and is able to detect very small amounts of synthetically concealed smuggled contraband.

Finally, a method for ATD is proposed, based on a deep Convolutional Neural Network (CNN), trained from scratch using TIP, and exploits the material information encoded within dual-energy X-ray images to suppress false alarms. The system offers a 100-fold improvement in the false positive rate over prior work.

When he took time to help the man up the mountain,
lo,
he scaled it himself

— Tibetan proverb

ACKNOWLEDGEMENTS

My sincere thanks to Dr Lewis Griffin, for his enthusiasm, understanding, support and guidance throughout this PhD, whilst affording me freedom and responsibility. And sincere thanks also to Dr Nicolas Jaccard, who acted as an excellent model to follow, whilst providing much insight, advice, and help. Thanks to both of you for making my PhD extremely agreeable.

Thanks, also, to all of the people that contributed to this research, including: Dr James Ollier who collected data for the wobble correction work, and provided good advice during the early days; Dr Edward Morton for whom the funding for this PhD would not have been possible; Dr Nick Calvert who helped me settle into UCL in the early days and provided much expertise on X-ray imaging. And to Dr Manos Protonotarios, who was willing to visit Stoke-on-Trent and help haul cargo in and out of a container on those cold and rainy mid-November days.

Thanks to everyone else in COMPASS, GriffinLab, SECReT and the Luncheon group, for everything.

Finally, thanks to my second supervisor Dr James Nelson (RIP), who was always enthusiastic and offered a unique perspective on this project.

This project was funded by Rapiscan Systems Ltd., the EPSRC, the UK Home Office, and Department for Transport. All cargo images were provided by Rapiscan Systems, and all baggage images by the UK Home Office unless stated otherwise.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AAD	Automated Anomaly Detection
ACV	Automated Contents Verification
AHE	Adaptive Histogram Equalisation
AQUA	Automatic QUality Assessment
AR	Auto-Regression
ATD	Automated Threat Detection
AUC	Area Under the Curve
BIF	Basic Image Feature
BM3D	Block Matching 3D
BPD	Beam Position Detector
BoW	Bag-of-Words
CAD	Computer Aided Design
CBS	Cabin Baggage Screening
CBT	Computer-Based Training
CNN	Convolutional Neural Network

CSI	Container Security Initiative
CSPIN	Colour SPIN
CSIFT	Colour SIFT
CT	Computed Tomography
DECT	Dual-Energy Computed Tomography
DEI	Dual-Energy Index
DEXA	Dual-Energy X-ray Absorptiometry
DGH	Density Gradient Histogram
DH	Density Histogram
DR	Detection Rate
DtG	Derivative of Gaussian
ECV	Empty Container Verification
EIP	Empty Image Projection
EM	Expectation-Maximisation
ESPIN	Energy SPIN
ETA	Euskadi Ta Askatasuna
FARC	Revolutionary Armed Forces of Colombia
FloodFill	Flood-Fill region growing
FNR	False Negative Rate
FPR	False Positive Rate
FTI	Fictional Threat Image
FRST	Forest of Random Split Trees
GMM	Gaussian Mixture Model
HHM	Hierarchical Holographic Model
HOG	Histogram of Oriented Gradients
HS	Harmonized System
ID	Imaging Detector
IDT	Isoperimetric Distance Tree
IED	Improvised Explosive Device
IRA	Irish Republican Army
LINAC	LINear ACcelerator
MAE	Mean Average Error
ML	Machine Learning
MV	Manifest Verification
MLP	Multi-Layer Perceptron
NII	Non-Intrusive Inspection
NIST	National Institute of Standards and Technology
NLM	Non-Local Means
oBIF	oriented Basic Image Feature
OCN	Organised Crime Network
OtF	On-the-Fly
PCA	Principal Component Analysis
PHOW	Pyramid Histograms of Visual Words
PSNR	Peak Signal-to-Noise Ratio
QE	Quantitative Evaluation
RBF-SVM	Radial Basis Function Support Vector Machine
ReLU	Rectified Linear Units
RANSAC	RANdom SAmple Consensus
RIFT	Rotation Invariant Feature Transform
RMS	Root-Mean-Square
RMSE	Root-Mean-Square Error
ROC	Receiver Operating Characteristic
ROI	Region-Of-Interest
RF	Random Forest
RRF	Random Regression Forest
RSS	Residual Sum of Squares
SIFT	Scale-Invariant Feature Transform

SURF	Speeded-Up Robust Features
SMT	Small Metallic Threat
SNR	Signal-to-Noise Ratio
SoC	Stream-of-Commerce
SSE	Sum of Squared Errors
SPIN	domain SPIN image descriptor
StD	Spot-the-Difference
SVM	Support Vector Machine
SymRG	Symmetric Region Growing
SI	International System of Units
TEU	Twenty-foot Equivalent Unit
TIP	Threat Image Projection
TIWS	Translation Invariant Wavelet Shrinkage
TPR	True Positive Rate
TV	Total Variation
USD	US Dollars
WMDs	Weapons of Mass Destruction

1

INTRODUCTION

EACH year, almost 7×10^8 Twenty-foot Equivalent Units (TEUs) of cargo container transactions occur globally [1], with a total approximate worth of 30 trillion US Dollars (USD) [2]. Assuming that each container is half full (the majority are fully packed), this amounts to a total of $2.4 \times 10^{10}\,\mathrm{m}^3$ of traded goods. And that is just for cargo containers; the number grows larger if bulk carriers, vehicles and other vessels are included. This is an impossibly large amount of goods to be physically inspected by humans, and it continues to grow each year [1]. Thus, the search for contraband is like searching for needles in an ever-growing field of haystacks (Figure 1.1).

The TEU is a common unit for measuring cargo volume. Each TEU is equivalent in volume to a standard 20 ft ISO container. Thus, a full 40 ft ISO container is equivalent to two TEUs.

Do any of these containers contain smuggled contraband?

Figure 1.1: An illustration of the task that a human operator is faced with - a stream of cargo container thumbnails. Seven of the loaded containers contain threats, and two of the 'empty' containers actually contain smuggled loads. In this thesis, automated methods are developed which are capable of handling this task.

As such, customs and border agencies have to prioritise which cargoes to inspect based on a risk analysis or specific intelligence from policing operations. The vast majority of cargo is not inspected at all [3]. This lack of surveillance is regularly exploited by Organised Crime Networks (OCNs), terrorists, and even 'legitimate' companies. It is surprising how, to date, research on automating cargo inspection has been so limited. Most work has been kept in-house at large security scanner manufacturers, and there has been little headway in the academic arena. This is in contrast to image analysis in

aviation security, where there has been far more government funding which has in turn stimulated academic research in the area. This is probably due to the more perceivable and imminent threat from terrorism, particularly in the wake of the 9/11 attacks.

This thesis is motivated by the current lack of research in automated cargo security, and the scale and importance of the task to the world economy and security.

1.1 DISCIPLINARITY

The work herein, first and foremost, proposes engineering solutions to cargo security problems. The work spreads a diverse range of subject areas including, most prominently; X-ray physics, image processing, machine learning, computer vision, and security science. Thus, it meets the multi-disciplinarity requirements of the UCL SECReT doctoral training centre.

1.2 CURRENT STATE OF THE FIELD

Research into automated analysis of X-ray cargo imagery is scarce, especially in comparison to passenger baggage in aviation security. This is due to the difficulty of obtaining suitably large datasets and the propensity of security scanner manufacturers to keep research private. However, from the body of work available, it is apparent that automated image analysis in cargo is following a similar development pattern to mainstream machine learning and computer vision.

The state of the field and how this thesis fits in is discussed in depth in Chapter 4.

Initial research proposed algorithms that were rule-based, in that threats were identified based on a hard-coding of the researcher's intuition or materials were classified based on methods derived from the physics of X-rays [4, 5]. Recently, methods have been developed that use hand-crafted features and Machine Learning (ML) approaches to combine features and to form a decision about the cargo container [6]. This thesis, begins with this latter approach and then aims to push the field forward through the use of end-to-end Deep Learning of features and their optimum combination scheme to realise automated image inspection.

1.3 CONTRIBUTIONS

This thesis makes the following novel contributions.

LITERATURE REVIEW The first review of the field of Automated X-ray Image Analysis for Cargo Security. Due to the relevant infancy of the field, no previous review has been published in the literature. The review covers all aspects of image analysis, including *image pre-processing* and *image understanding*. The published version of the review [7] is longer than the version presented herein, with more analysis on the future direction and promise of the field.

WOBBLE MEASUREMENT A novel method of measuring detector wobble in large-scale transmission radiography is developed. Wobble is estimated based on measurements from four imaging detectors rotated by 90° to measure the profile of the X-ray beam across its width. In this thesis, these rotated detectors are referred to as Beam Position Detectors (BPDs). During a scan BPDs are obscured by arbitrary objects in the scene, and thus obtaining wobble estimates is difficult. The proposed solution uses a Random Regression Forest (RRF) model to obtain an instantaneous wobble estimate, and an Auto-Regression (AR) model to determine an estimate based on previous estimates. The RRF and AR estimates are then fused using a Bayesian approach to form a superior estimate of wobble. This is the first attempt, according to the literature, of measuring wobble in large-scale transmission radiography.

WOBBLE CORRECTION A novel method for correcting image intensity variation due to detector wobble is presented. The method is derived by considering a model of image formation in the presence of a wobbling detector. Wobble correction also relies on the estimation of misplacements and rotations of individual imaging detectors that can complicate wobble correction. This is the first attempt, according to the literature, of correcting wobble in large-scale transmission radiography. The correction method, when combined with the measurement method, is able to correct for 87% of image intensity variation due to wobble.

TIP, DATA AUGMENTATION, AND EIP A method for Threat Image Projection (TIP) in cargo is proposed and validated experimentally. TIP is typ-

ically used in aviation security to blend threat items into benign X-ray images, which are then used to train and evaluate human operators. Here, we propose that a similar approach can be used to train and evaluate threat detection algorithms. The TIP method relies on the Beer-Lambert law, and the experimental validation shows that there is no qualitative or quantitative difference between a real threat image and its equivalent TIP image. Methods of injecting natural variation into TIP imagery are proposed as a form of data augmentation, useful for training ML-based algorithms. Moreover, due to concerns that ML systems can learn subtle cues from artefacts present in TIP imagery, an Empty Image Projection (EIP) method is proposed that can be used in various ways to improve confidence that this is not occurring. The aim of EIP is to project just the TIP artefacts and not the actual threat.

EMPTY CONTAINER VERIFICATION The first ML-based method for Empty Container Verification (ECV) is presented. ECV classifies whether a cargo container is empty or non-empty, which is most useful for detecting contraband, fraud and errors within empty containers which constitute 20% of the global container fleet. The method works by Random Forest (RF) classification of individual image patches. The features used include fixed geometric features and the coordinates of the patch within the image. The use of window coordinates as a feature allows the RF model to learn the location-specific range of appearances of an empty container and overcomes the need to segment the image and apply a separate classifier to each image region. The method is trained on TIP-generated non-empty examples and real empty examples from the SoC. The proposed ECV method outperforms a previous method reported in the literature when tested on the SoC, and is also tested for very small TIP-generated adversarial loads that are representative of smuggled cocaine.

Stream-of-Commerce (SoC) containers are containers imaged in real-life, and not in a laboratory experiment.

The threats in question are censored to prevent keyword searching, but have recently been used to commit a range of so-called Mumbai-style terrorist attacks [8] and are explicitly considered in Roomi and Rajashankari [9].

DUAL-ENERGY AUTOMATED THREAT DETECTION A method for detecting Small Metallic Threats (SMTs) in complex dual-energy cargo imagery is presented. According to the literature, it is the first time in cargo that Automated Threat Detection (ATD) has explicitly operated on images measured at different energies. The motivation is that the algorithm can learn to suppress false positives by using material information derived from the dual-energy measurement. The method employs a trained-from-scratch Convolutional Neural Network (CNN), trained on TIP imagery with data augmentation, and

significantly improves on the prior work in the literature [10]. The system is also validated using EIP.

1.4 PUBLICATIONS

The work detailed in this thesis has directly contributed to the following publications:

- T. W. Rogers et al. (2016a). 'Automated X-ray image analysis for cargo security: Critical review and future promise'. *Journal of X-Ray Science and Technology* **25**.1, pp. 33–56.
- T. W. Rogers et al. (2014). 'Reduction of Wobble Artefacts in Images from Mobile Transmission X-ray Vehicle Scanners'. In: *Proc. IEEE Imaging Systems and Techniques*, pp. 356–360.
- T. W. Rogers et al. (2016b). 'Measuring and correcting wobble in large-scale transmission radiography'. *Journal of X-Ray Science and Technology* **25**.1, pp. 55–77.
- T. W. Rogers et al. (2016c). 'Threat Image Projection (TIP) into X-ray images of cargo containers for training humans and machines'. In: *Proc. IEEE International Carnahan Conference on Security Technology*, pp. 1–7.
- T. W. Rogers et al. (2015). 'Detection of cargo container loads from X-ray images'. In: *Proc. IET Intelligent Signal Processing*, pp. 1–6.
- T. W. Rogers et al. (2017). 'A deep learning framework for the automated inspection of complex dual-energy x-ray cargo imagery'. In: *Proc. SPIE*. **10187**, pp. 1–12.

Several other publications have been published during the PhD, which are not covered in this thesis:

- M Caldwell et al. (2017). 'Transferring x-ray based automated threat detection between scanners with different energies and resolutions'. In: *Proc. SPIE*. **10441**, p. 1
- N. Jaccard et al. (2016c). 'Tackling the X-ray cargo inspection challenge using machine learning'. In: *Proc. SPIE*. **9847**, pp. 1–13.
- N. Jaccard et al. (2016a). 'Automated detection of smuggled high-risk security threats using Deep Learning'. In: *Proc. IET Imaging for Crime Detection and Prevention*, pp. 11–15

- N. Jaccard et al. (2015). 'Using deep learning on X-ray images to detect threats'. In: *Proc. Cranfield Defence and Security Doctoral Symposium*, pp. 1–12.
- N. Jaccard et al. (2016b). 'Detection of concealed cars in complex cargo X-ray imagery using deep learning'. *Journal of X-ray Science and Technology* **25**.3, pp. 323–339.
- N. Jaccard et al. (2014). 'Automated detection of cars in transmission X-ray images of freight containers'. In: *Proc. IEEE Advanced Video and Signal Based Surveillance*, pp. 387–392.
- J. T. A. Andrews et al. (2017). 'Representation-learning for anomaly detection in complex x-ray cargo imagery'. In: *Proc. SPIE*. **10187**, pp. 1–11.
- J. T. A. Andrews et al. (2016a). 'Anomaly Detection for Security Imaging'. In: *Proc. Cranfield Defence and Security Doctoral Symposium*, pp. 1–14.

Some of this work has been presented at the House of Commons, as part of the SET for BRITAIN competition, where it was awarded Silver and ranked top out of all Engineering PhD projects. Recently the Economist [23] featured an article on the ATD work.

Some passages and figures in this thesis have been quoted verbatim, with permission, from Refs. [7, 12–14]. All text and figures originated from myself, unless indicated.

1.5 THESIS STRUCTURE

The context for the technical work in this thesis is set out in Chapter 2, through a taxonomy of the different ways in which the containers that compose the global supply chain are exploited by OCNs and terrorists. In addition, an overview is provided of the current security infrastructure in place to help detect and deter such activity, and how the technologies developed in this thesis can fit into this infrastructure. This chapter also serves as an extended motivation for this thesis.

The fundamentals of large-scale transmission X-ray imaging are presented in Chapter 3. This includes: the fundamentals of X-ray generation, interactions with matter, and detection; the practical deployment of cargo X-ray

scanners; and the typical properties, noise and artefacts of x-ray cargo imagery. It provides a physics underpinning for the technical work in this thesis.

A review of the literature on automated cargo image analysis is presented and dissected in Chapter 4. The review covers the themes of *image preprocessing* and *image understanding*. It ends with a discussion on how this thesis, and other PhD publications not presented herein, fit into the existing literature.

In Chapter 5, detector wobble and its effect on image quality are described. A model of image formation in the presence of a wobbling detector is deduced, and from this a number of image corrections are derived. Next, a ML-based method for the measurement of wobble is proposed and tested. Finally, the wobble estimation and image correction methods are combined and tested on wobble-effected images.

Chapter 6 introduces a framework for the synthesis of threat imagery by TIP. The TIP framework can equally be used to train humans and machines. The method is validated experimentally. In addition, the concept of EIP is introduced, and a method proposed to implement it. Both TIP and EIP are used for training, evaluating and probing ML-based *image understanding* in the next two chapters.

Chapter 7, the first on *image understanding*, proposes a method for automated ECV. The method is tested on both real images of cargo containers from the SoC, and against small adversarial loads generated using TIP. TIP is also used to better characterise the performance of the system as a function of the location within the container, and the volume and density of the adversarial load.

Chapter 8, the final technical contribution of this work, addresses the detection of SMTs in complex dual-energy X-ray images. Several dual-energy CNN architectures are trained from scratch, using TIP and data augmentation, and tested on TIP imagery. The method is also validated using EIP to ensure that it is not gaining a major advantage over real threat data by learning to exploit subtle artefactual cues from the TIP process.

Finally, a critical review and discussion of the research contributions of the preceding chapters is given in Chapter 9, as well as proposals for future directions of research.

2

SUPPLY CHAIN CRIME AND SECURITY

This chapter lays out the real world context of this research, describing the scale and importance of the global supply chain and how Organised Crime Networks (OCNs), terrorists, and even legitimate businesses seek to exploit its vulnerabilities. The cargo crime and security landscape is quite complex, and this chapter aims to distil it into a taxonomy of its main elements.

2.1 THE GLOBAL SUPPLY CHAIN

The global supply chain is a complex network of suppliers, factories, warehouses, distribution centres, ports and retail outlets [24]. The supply chain converts raw materials into finished products which are eventually delivered to the consumer. Over time, the supply chain has evolved to become increasingly efficient and resilient, as stakeholders have sought to extract greater profits and reduce risk [24].

The major breakthrough in network efficiency came in the mid-twentieth century, with the introduction of standardised and discretised units; intermodal containers. These units, originally developed by the US military, can be carried across different modes of transport in the supply chain; by ship, rail, or truck [25]. Intermodal containers have the benefit that they can be stacked efficiently (Figure 2.1) in large numbers at ports or on container ships, and carried individually on land by truck or rail. The standardised system of containers revolutionised the supply chain, and today, 90% of world cargo is shipped in containers [26]. Since the inception of intermodal containers, stakeholders have invested large sums of money on optimising container routing and security. Initially, most security investments focused on reducing the number of containers lost through theft and misrouting, as this has direct commercial implications on stakeholders in the private sector [24].

Intermodal containers go by a number of names, including: ISO container; shipping, sea or ocean container; cargo or freight container. These are used interchangeably throughout this thesis.

After the 9/11 terrorist attacks in the US, the attention of security practitioners shifted onto the risks that containerisation can pose. The global

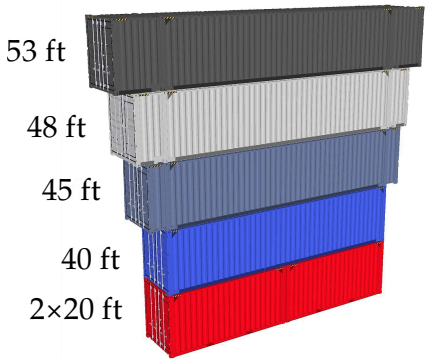


Figure 2.1: Illustration of a stack of intermodal containers of increasing volume. Cargo containers range in size from 10 ft to 60 ft, whilst 90% of all containers are either 20 ft or 40 ft long. The ability to stack discretised units at ports or on ships, or carry them individually by rail or road has revolutionised the efficiency of the global supply chain.

supply chain is the backbone of global commerce and the system has sufficient redundancy to make it extremely unlikely that some event will grind it to a halt; if it did, the result would be a catastrophe for the world economy. However, localised attacks can still cause serious damage, particularly to the countries in which they occur. In 2002, the 10-day US West Coast port lockout caused economic damage worth between 4.7 and 19 billion US Dollars (USD) [27, 28]. Particularly in the US, governments have been concerned about attacks or hoax attacks bringing ports to a standstill and causing untold economic damage [24].

Governments have invested significant sums to develop Non-Intrusive Inspection (NII) techniques to detect tell-tale threat signatures or to visualise an image of the cargo contents. Most research initially focused on the detection of Weapons of Mass Destruction (WMDs) or their precursors [24]. Particular attention was paid to the detection of nuclear materials through radiation monitoring. Over time, the NII methods have improved enough to allow customs and border officers to search containers for contraband or shipping fraud, with transmission X-ray scanners the most common [29].

A WMD is any nuclear, radiological, biological or other weapon that can cause mass loss of life or injury, or significant damage to man-made or natural structures, or the biosphere.

From 2004 to 2014, and despite the 2008 global economic crisis, the number of Twenty-foot Equivalent Unit (TEU) transactions more than doubled to reach almost 7×10^8 TEU per annum [1]. During this time, the US Container Security Initiative (CSI), proposed in the wake of the 9/11 terrorist attacks, has encouraged 100% screening of containers [30], but has yet to be fully realised [31]. With the increasing number of containers and the increased screening requirements, providing high levels of cargo security is becoming evermore difficult.

2.2 A TAXONOMY OF CARGO CRIME

As legitimate businesses have invested money to improve the efficiency of the supply chain, OCNs have invested money to learn how to best exploit it. OCNs are similar to a business in that they seek to maximise benefits and minimise costs. Indeed, OCNs can be modelled using the rational choice theory of economics, where actors make decisions based on a cost-benefit analysis [33]. And like a business, an OCN has a product to sell, and a lucrative market to sell it to. To get their product into international markets, the OCN must use logistics. Based on this, it is evident that criminals would seek to exploit the already highly optimised global supply network and the intermodal containers that constitute it. There are even OCNs that work solely on providing logistics to other OCNs by establishing safe smuggling routes [34].

There is no legal definition of organised crime in the UK, although it is considered as 'serious crime planned, coordinated and conducted by people working together on a continuing bases' in the UK Serious and Organised Crime Strategy [32].

A Hierarchical Holographic Model (HHM) [35] taxonomy of the *modus operandi* used in cargo crime is given in Figure 2.2. OCNs use a range of sophisticated methods, often in tandem, when committing crimes. These include:

(i) Smuggling - the act of physically concealing smuggled contraband within the container.

(ii) Manifest fraud - the act of falsifying container manifests to facilitate smuggling.

(iii) Container routing fraud - the act of concealing the shipment route from authorities to facilitate smuggling.

(iv) Use of agents - the use of human resources, often people involved with enforcing cargo security, and who can be manipulated to facilitate smuglling.

The goal of these crimes is to facilitate the criminal logistics of illicit products.

2.2.1 *Smuggling*

Smuggling bypasses customs controls, allowing criminals to: avoid duties on legitimate goods (e.g. cars, alcohol, cigarettes); trade prohibited or counterfeit items; or transport prohibited items such as stowaways, counterfeit goods, animals, and stolen goods [37]. A taxonomy of contraband examples

As of 2007, it was estimated that 1.95%, by value, of world trade was of counterfeit goods. Counterfeit items were worth an estimated 2.5×10^{11} USD [36].

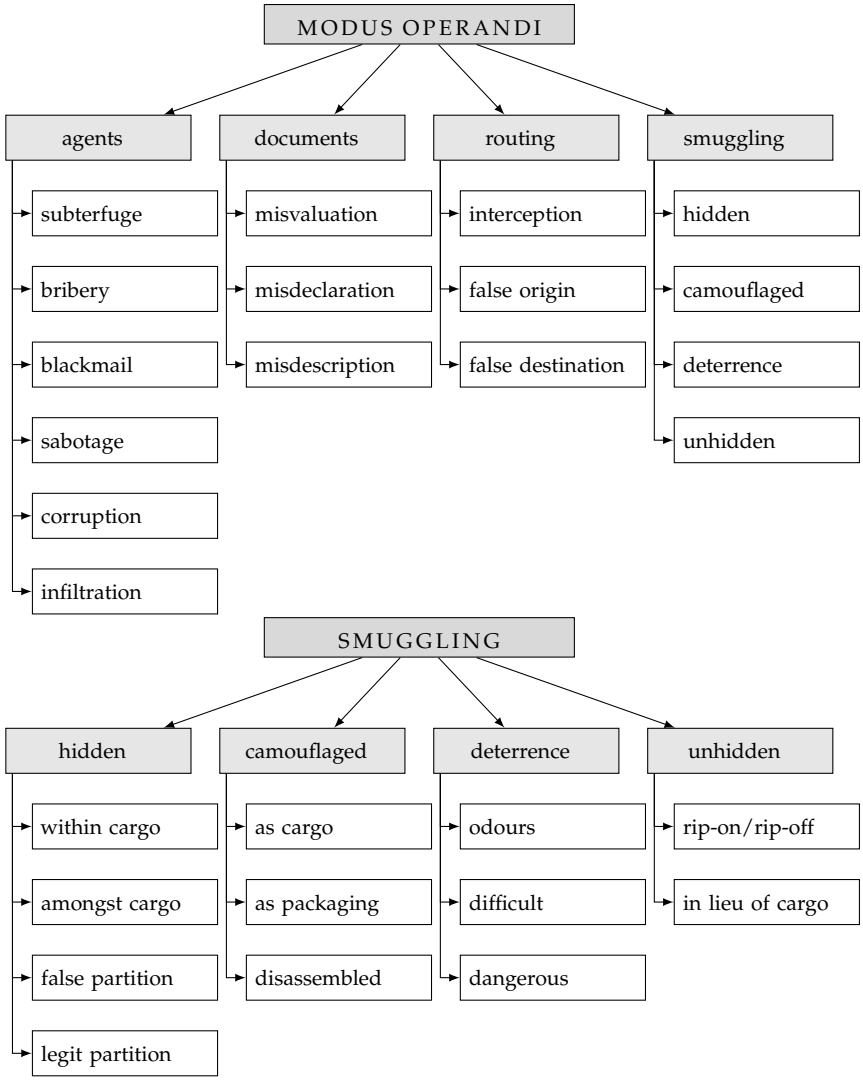


Figure 2.2: Hierarchical Holographic Model (HHM) of the *modus operandi* of cargo container smugglers.

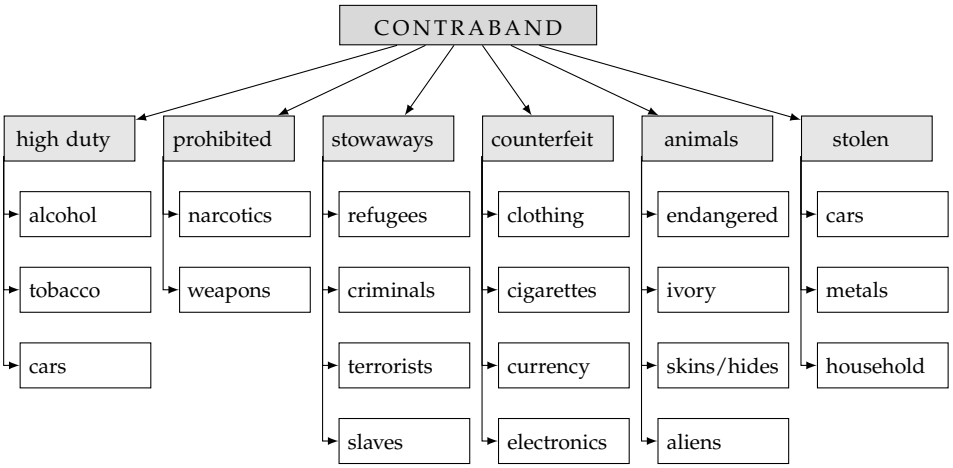


Figure 2.3: Hierarchical Holographic Model (HHM) of smuggled contraband.

is given in Figure 2.3. Successful OCNs take a scientific approach to smuggling; they will invent new techniques, test and refine them to find the ones that bring most success. Often a team of 'surgeons' is employed for this purpose [38].

The smuggling techniques range in sophistication, aiming for a balance between the cost of the smuggling technique, the likelihood of the goods being seized, and the potential cost to the OCN if the seizure occurs. The least sophisticated methods hide contraband in plain sight (unhidden) and hope that the container is not inspected at all. With increasing cost, methods may conceal contraband amongst legitimate cargo, or within individual items of legitimate cargo. Tactics may also be used to deter customs officers from physical inspection.

For lack of a better word, we use 'unhidden'.

(a) 1×10^6 Cigarettes (b) 514 kg Cocaine

Figure 2.4: Examples of concealment within a legitimate (a) and a false partition (b). Source: EU Commission [39].

HIDDEN CONTRABAND Contraband can be hidden within container partitions, and amongst or within legitimate cargo. Container partitions can be legitimate or false. Examples of legitimate partitions include the container cross-beams, the refrigeration unit in refrigerated containers, or compartments in the container roof and floor (Figure 2.4a). OCNs retrofit false partitions to the container. Typically false walls or bulkheads (Figure 2.4b) are installed, sometimes with a hidden entrance on the top of the container so that it can be loaded with contraband. Even declared-as-empty containers have been found with contraband stashed in false and legitimate partitions [39]. Contraband may also be hidden within legitimate items (Figure 2.5c), or amongst legitimate cargo (Figures 2.5a, 2.5b and 2.5d).

(a) Appears to be household goods

(b) A stolen car is revealed

(c) Counterfeit cigarettes hidden in sofa

(d) Missile hidden in pickled cucumbers

Figure 2.5: Examples of concealment within legitimate cargo shipments. Sources: UK National Crime Agency and EU Commission [39].

CAMOUFLAGED CONTRABAND One of the most sophisticated smuggling *modus operandi* is to disguise contraband as legitimate items. Cocaine is often mixed with liquid cosmetics or oil [38]. There have been seizures where cocaine has been mixed into plastic during the manufacturing process of replica commercial products. In these cases the cocaine can be very difficult to detect, particularly when mixed with plastics, even during physical inspection. However, it is unclear how efficiently the contraband can be recovered. In addition to visual camouflage, substances such as naphthalene may be used to disguise the scent of narcotics.

DETERRENCE Often, OCNs experiment with new ways to deter customs officers from physical inspection. This can include the use of unpleasant odours such as excrement, making the cargo difficult or laborious to unpack by planting heavy loads, or by placing items in the container that are dangerous if touched [40]. However, it is unclear whether such tactics work because they can often arouse the suspicions of the inspector. Figure 2.6a shows an example of where criminals experimented with using a beehive to deter inspectors.

UNHIDDEN CONTRABAND If there is high confidence that a shipment will not be inspected, OCNs can save money on the costs of hiding contra-

(a) A beehive deterrent

(b) Rip-on/rip-off bag

Figure 2.6: Examples of deterrence and unhidden contraband. The beehive was used to deter border officers from inspecting the box which contains cocaine bricks and trousers. The rip-on/rip-off bag contains 40 kg cocaine, and attached is a fresh seal (inset) so that the container can be resealed after extraction. Source: EU Commission [39].

band. For example, they may have used agents to ensure that the shipment is not inspected, or falsified manifest information to make the shipment appear 'low risk'. Another reason to leave contraband unhidden is so that it can easily be planted and extracted from the container in what is known as a rip-on/rip-off attack. This relies on breaking into the container during transit, when it is least likely to be inspected. It is essential that this is done fast to reduce the risk of getting caught, and this is possible if contraband is left unhidden. In rip-on/rip-off attacks, contraband is usually stored in a sports bag so that it can be easily carried away (Figure 2.6b). Since 20% of all containers are declared-as-empty and often left unlocked, they are particularly vulnerable to rip-on/rip-off attacks.

This thesis aims to address this issue, as well as hidden contraband in otherwise empty containers, in Chapter 7.

2.2.2 *Use of agents*

In addition to the method of concealment, OCNs employ deception, inside agents, sabotage, bribery, blackmail, and corruption. Without too much imagination, a number of fictional scenarios can be constructed based on global experience in the area [34], as listed below.

DECEPTION A lorry driver is asked to smuggle duty-free cigarettes in his lorry. With the monetary incentive offered by his contact, and believing that smuggling duty-free cigarettes is not too unethical because they are too highly taxed, he agrees. After a year of smuggling cigarettes, they are swapped for heroin without him knowing, and so he inadvertently smuggles huge volumes of a Class A narcotic, something he would never have agreed to originally.

INSIDE AGENTS A UK port hires an ex-convict as a cleaner. After a few years, the ex-convict is promoted and trained to operate cranes, but is not allowed to inspect cargo due to his criminal history. The ex-convict moves a cargo container to a quiet part of the port where it is forgotten about and avoids inspection. A few months later the container is found and opened by another port worker and it is found to contain 30 stowaway cadavers. The container was supposed to be picked up by truck, but it was found too risky due to the surveillance at the port.

BLACKMAIL A port worker has been monitored for several months by an OCN and found to be using prostitutes. The OCN threaten to tell his employers and family, unless he does a job for them. The job is simply to open up a container, retrieve a locked metal case and to take it to a public toilet at 20:00. The case contains two AK47s, ammunition, and grenades. A month later, there is a Mumbai-style terrorist attack in London, with 61 people killed. The weapons were purchased by the terrorists on the Deep Web.

BRIBERY A port worker is in financial difficulty with several unpaid credit card debts, and he needs to replace his broken car. He meets a person in the pub who says he will pay him £10,000 to falsify a container inspection. The container is from Columbia and contains shoes as well as 100 kg of cocaine with a street value of £4,600,000. The next day there are reports that a truck has been hijacked and the driver missing. The port worker is found dead in his basement the next day, and believed to have taken his own life. The £10,000 was never paid and the worker was actually murdered.

2.2.3 *Document fraud*

Harmonized System (HS) is also known as the Harmonised Commodity Description and Coding System.

Each cargo container has an associated manifest document that describes broadly what is contained, including the mass, value, and types of cargo. The type of cargo is characterised by a HS code. The codes are hierarchical. For example, the HS code for a typical wine is constructed as:

$$\underline{22} \implies \text{Beverages, spirits and vinegar}$$
$$22\underline{04} \implies \text{Wine of fresh grapes, including fortified wines}$$
$$2204.\underline{21} \implies \text{In containers holding 2 litres or less}$$
$$2204.21.\underline{4210} \implies \text{Of alcoholic strength by volume} \leqslant 13\%.$$

Criminals, individuals and businesses can intentionally falsify manifests in order to avoid tariffs and duties. The types of falsification include [37]: false HS codes (tariff); undervaluation; over-valuation; under-declaration; and over-declaration. Such methods can also help evade import restrictions and exchange controls, and enable false claims for export refunds, remissions and drawbacks. Moreover, false licenses and permits may be used to help evade quotas and prohibitions. These can be obtained by theft, unauthorised completion of genuine blank documents, or forgery and alteration of existing documents [37].

Finally, the recorded cargo origin and destination can be falsified, giving entitlement to preferential duty rates for products from specified countries, and allowing evasion of quota restrictions or anti-dumping regulations applied to certain countries. Falsifying the destination can allow evasion of export restrictions to particular countries. This can be achieved by shipping the cargo to a middleman, who forwards the consignment to the actual destination.

2.3 SPILLOVER CRIME

Apart from the crimes listed in the taxonomy (Figure 2.2), cargo crime can lead on to other forms of crime, in what is known as 'spillover crime' [41]. For example, smuggling brings in revenue for the OCN, which can be reinvested into other illegal activities. Moreover, the revenue from smuggling helps the OCN survive, if not grow, and OCNs are heavily associated with violence.

For OCNs violence is vital; it provides both a means for internal control and a weapon that is highly visible from the outside [42]. Often, unlike a legitimate business, OCNs cannot use the law to settle disputes with rivals, and so have to resort to violent turf war. OCNs are also highly paranoid that they have been infiltrated by law enforcement, or rival networks [42]. And

since an OCN operates to reduce risk, if they have the slightest suspicion about a member, then they take the lowest risk solution; murder.

This is often referred to as the '9 mm rule' by law enforcement [34].

The proceeds from smuggling are often laundered and concealed within front business activities. OCNs have to avoid the scrutiny of bank regulators and investigators, which requires them to have in-house specialists including financial advisers, accountants, and bankers [42]. Often bribery and blackmail is used to hire specialists, however they can also be deceived, or fully cognisant but motivated by a large salary. Even bankers in major financial centres have been found to work for OCNs [42].

By targeting and disrupting the logistics of OCNs, one can harm their profits, and in turn reduce spillover crime.

2.4 TERRORISM

There is a nexus between organised crime and terrorism [43].

Firstly, for terrorists to execute attacks they need to obtain a weapon. Typically, home-made explosives were used, however, there has been recent inspiration from the 2008 Mumbai attacks where terrorists used firearms indiscriminately on soft targets. In recent years, the number of Mumbai-style attacks [8] have increased as evidenced by the attacks on Paris [44] and on the beaches of Tunisia [45]. It is argued that most of these weapons come from the Balkan states, which are awash with illegal firearms left over from conflicts in the 1990s [43, 46]. So there is a direct connection between organised arms traffickers (a type of OCN) in the Balkans and terrorist attacks committed elsewhere in Europe. The firearms may not intentionally be sold to the terrorists, however the trafficking increases the pool of weapons within a country, giving terrorists more opportunity to obtain one.

In fact, opportunity is seen by many crime theories as the key driver for crime [47, 48]. A would-be terrorist might know someone with a weapon and purchase, borrow, or steal it.

There is also a link between terrorism and OCNs specialising in narcotics trafficking [43], due to the borderless aspect of crime. For example, the Revolutionary Armed Forces of Colombia (FARC), whose annual income from narcotics is as high as 3 billion USD per annum [49], often make transactions with the Irish Republican Army (IRA) in narcotics. In return, the IRA provides the FARC with training in terrorist and guerilla techniques [43]. The IRA has been found to trade narcotics with Croatian arms traffickers. Another example is the Euskadi Ta Askatasuna (ETA) who have also used narcotics as barter for weapons from the Balkans to support their terrorist endeavours [43]. The ETA have also been known to receive heavy weapons

from the Camorra OCN in Italy, and to have provided explosives to Hamas in Palestine [43]. In Northern Ireland, it was found that over half of gangs have links to republican or loyalist paramilitary groups, and it is estimated that 66% of gang members are involved in narcotics, 55% in counterfeiting and 50% in money laundering [43].

Due to this complex nexus between OCNs and terrorism, a robust approach to reducing the risk of terrorism is to target and disrupt all types of organised crime, including cargo smuggling, and not just the movements of terrorist implements.

2.5 CONTAINER SCREENING

Cargo screening is performed in three parts: (i) selection of containers for inspection; (ii) NII of the container contents; and (iii) physical inspection of the container and its contents. These components increase in cost, and so nearly all containers are passed through selection, few of these are selected for NII, and even fewer are physically inspected.

Some figures suggest only 4-5% of containers currently undergo NII [3].

2.5.1 *Container selection*

Containers are selected based on a risk analysis, specific intelligence from ongoing criminal investigations, or at random. The container risk is often assessed by an automated system based on the shipping manifest information. Risk indicators can be associated with [40]:

- *Involved parties* – absence of cooperation from the economic operator, use of a carrier with dubious reliability, or involvement of suspicious companies;
- *Involved goods* – goods systematically labelled as 'others', dubious or absent labelling, discordance between the weight and number of packages, and inadequate packaging for the type of goods;
- *Shipment route* – unjustified route involving transshipment, or an uneconomic route.

Companies may be deemed suspicious if they; are newly-formed, have a weak financial structure, have previous history of fraud, or if the shipment originates from a known high risk geographical region [40].

Since the risk analysis is usually informed by prior seizures, criminals may purposefully use low risk countries or manifest details. Therefore it is important to randomly select a fraction of low risk containers for inspection. New discoveries can be used to update the risk analysis.

2.5.2 *Non-intrusive inspection*

There are a wide range of image-based NII technologies available to customs officers, which fall into two general categories; active and passive radiography. In active radiography, the container is probed using a radiographic source and detectors. The main types of active radiation used are γ-rays, X-rays, and neutrons. Passive radiography exploits natural radiation sources, such as neutron or γ radiation emitted from goods inside the container or muons from cosmic radiation.

Of the active radiation systems, transmission γ- and X-ray radiography are the most common [29]. Such systems are well developed and can come in a number of different configurations, including multiple views to make apparent heavily obscured objects, and multiple source energies to help discriminate between different materials.

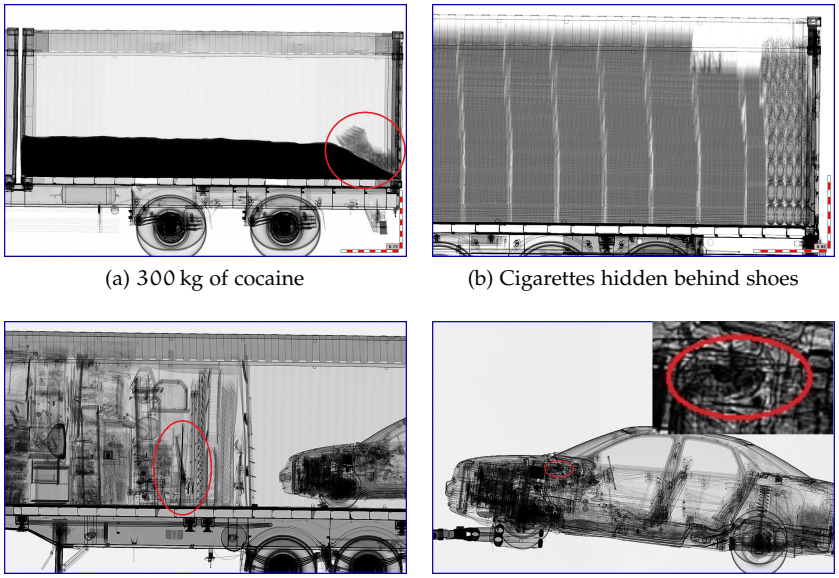
In this thesis single-view, and single- and dual-energy transmission X-ray scanners are employed. These systems are described in detail in Chapter 3.

The type of photon sources differ between γ- and X-ray based systems; X-rays are usually created as Bremsstrahlung radiation from colliding accelerated electrons into a metal target, and γ-rays are emitted from a radioactive source. X-ray cargo systems operate at higher energies than γ-ray cargo systems, thus offering greater penetration. However, γ-ray sources emit photons at discrete energies making material discrimination more straightforward, in theory, than when using continuous Bremsstrahlung spectra.

With multiple views it is possible to reconstruct 3D information about the container and contents [50], however due to the size of containers and the number of views needed to reconstruct an accurate 3D image, these systems are not in wide use. More recently, backscatter X-ray systems have been developed, which operate at lower energies, and measure X-rays that are scattered back towards the source. Backscatter systems are useful for the detection of low atomic number signatures, such as narcotics and cigarettes. Research is ongoing into the possibilities of forming 3D images by measuring the time-of-flight of individual X-rays [51].

Passive imaging systems are less common for cargo screening, although non-imaging systems such as radiation monitors are commonplace. Neutron-based systems can be designed to exploit different neutron interactions including scatter and thermal neutron capture. The benefits of these systems over X-ray, is that they give a much greater degree of material discrimination, particularly for organic materials. However, neutron-based systems offer less penetration than X-ray or γ-ray systems [29]. Another passive tech-

nology is muon tomography, which uses natural muons from cosmic rays. By detecting the momentum and trajectory of muons, a 3D image of the container contents can be reconstructed and it is possible to achieve material discrimination [52]. The technology is still in its infancy, but offers some promise.



(a) 300 kg of cocaine

(b) Cigarettes hidden behind shoes

(c) Hidden hunting rifles

(d) Handgun

Figure 2.7: X-ray images of smuggled contraband. Images courtesy of Dutch Customs.

The captured image is inspected by a customs officer, who generally looks for two things when detecting threats; tell-tale silhouettes and anomalies [53]. When searching for tell-tale silhouettes, operators search for objects in the image that are similar in size, shape and density to known threats. For example, weapons and military equipment have a distinctive silhouette in comparison to common legitimate items. Anomaly detection is useful if the image is very confusing due to clutter, the threat is very small, or if the adversary has masked the threat with dense materials. Anomalies can be indicated by breaks in symmetry, or changes in density or texture. Additionally operators may look for wavy or turbulent patterns which may indicate illegal liquids sloshing around [53]. Figure 2.7c gives an example where a weapon can be found by searching for a tell-tale silhouette and Figure 2.7d by looking for an anomaly. The narcotics and cigarettes in Figures 2.7a and 2.7b can be found either by looking for silhouettes or anomalies.

In addition to searching for specific threats or anomalies, customs officers also compare the image appearance to the manifest document. This is known as Manifest Verification (MV). The simplest form of MV is to check whether a declared-as-empty container is actually empty and does not contain smuggled contraband, or has been declared-as-empty to avoid tariffs on legitimate items. Moreover, more advanced MV could include estimating the HS code types, numbers and masses of objects in the container and comparing with the manifest.

If a threat, anomaly, or manifest discrepancy are discovered in the image, the container is sent for physical inspection.

2.5.3 *Physical inspection*

Physical inspection is a long and arduous process, and can lead to shipment delays of between a day and several weeks [54]. Before the container is even opened, the exterior of the container is carefully checked. Particular attention is paid to the locks and seals, signs of container repair and tampering, new paintwork or mastics, and alterations to the container numbering [40]. Sniffer dogs are often used to scout out the container exterior for drugs or explosives. Sometimes, if the container is padlocked, a locksmith is required to gain access. This process is all carefully documented and photographed in case it is needed as evidence [40].

Next, officers open the container carefully checking that the cargo is not leaning against the door. The condition and arrangement of items in the container is carefully documented, in case there is a later claim for damages. Heed is given to the warning signs (e.g. flammable, toxic) placed on the goods to ensure the proper precautions are taken. Officers pay attention to unusual odours such as excrement or naphthalene, which can be used to camouflage drug odours or deter officers [40]. If possible all of the goods are removed from the container and undergo individual inspection. The officers will compare the weights of 'identical' packages, and inspect them for visual irregularities, such as damage or glue that might indicate tampering [40]. Officers often use hand-held devices to non-intrusively test individual packages.

After goods inspection, the container interior is checked for signs of false partitions or tampered-with legitimate partitions. For example, a tape measure can be used to check the interior dimensions, if they are shorter than

expected then this indicates a possible false partition [40]. Some containers are more difficult to inspect, such a refrigeration units where the insulation should be removed and inspected for hidden contraband [40].

When a smuggling attempt is discovered, the customs officers consult with law enforcement on the next actions. Sometimes it is best to leave the contraband in place, or substitute it for dummy contraband, and keep the container under surveillance [37]. This can help the authorities gather more intelligence on the OCN or terrorist network, helping them inflict maximum damage to the network. It is worth noting that not all smuggling attempts have to be detected, since an OCN may make many attempts, and detecting just one of them can take out large parts of the network after further investigations. At the same time, it is important to limit false positive searches since physical inspection is expensive and time-consuming.

2.6 THE FUTURE OF SUPPLY CHAIN AND BORDER SECURITY

Due to the importance of a high throughput to the world economy, investments are being made to automate parts of the supply chain. In the UK, a recent example is London Gateway, which uses fully automated cranes and traditional artificial intelligence task planning and optimisation strategies to pick and stack cargo containers, as well as route trucks to the correct region of the port to pick up a container [55]. In the future, to cope with increasing throughput demands, automation is likely to stretch to security too. Indeed, the UK Government has recently announced plans to increase the amount of automation in its National Security Strategy and Strategic Defence and Security Review [56]. In particular, the government is concerned about the illegal movements of firearms and their potential use in terrorist attacks and by OCNs:

> *'We will strengthen our ability to detect the movements of people and goods such as illegal firearms that present a threat, through detection technology and better data on land, sea and air passengers and cargo, and through intelligence and targeting. We will modernise and introduce more automation to enable us to deploy our border officers where they are needed most, including an enhanced presence at our sea ports.'* — HM GOVERNMENT [56]

At present, there is a compromise between security and throughput. High security, without automation, inevitably leads to a reduction in throughput. This is particularly evident in airports, where all passengers and baggage

are screened, but this takes time, leading to queues and frustrated passengers. Cargo is currently in the opposite paradigm. Throughput is typically very high, but security is weak because only a small fraction of selected containers are inspected even non-intrusively.

For cargo, there are a number of systems that can scan all containers automatically at high speed. For example the high-speed Rapiscan Eagle® R60, can scan containers carried by rail at 60 km/h and automatically detects when containers are passing through the scanner to form individual container images, and turns off when the driver's cabin is passing through. If such scanners were used at all ports for cargo entering or exiting the port by land, potentially every container can be imaged automatically. However, with images taking up to 15 minutes for a full inspection [57], the number of images is overwhelming for human operators. Automated security, through automated image analysis, can help to negate the security-throughput compromise. Particularly if automated inspection can be performed at a similar or better performance than humans in terms of accuracy and speed.

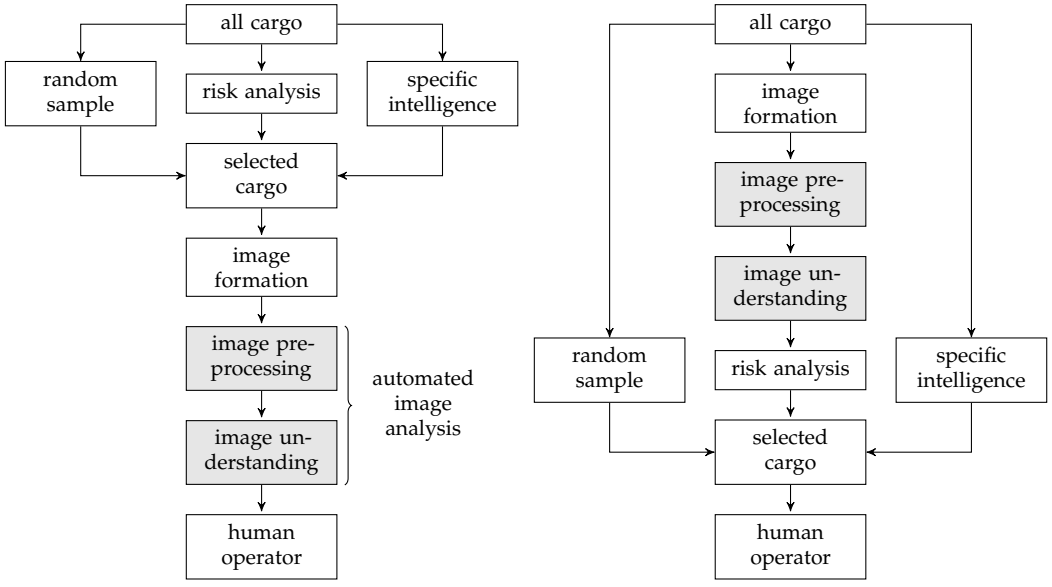


Figure 2.8: Process diagrams showing the cargo inspection process and possible uses of automated image analysis; Assisted Inspection, and Assisted Selection. In Assisted Inspection, annotations (e.g. bounding boxes with a labelled confidence score) determined in the Image Understanding step are added to the image so that the operator can more quickly identify threats. In Assisted Selection, image analysis is used to inform the risk analysis so that cargo is selected for inspection with greater accuracy, and thus reducing the burden on human operators.

Automated image analysis can help with cargo screening by Assisted Inspection or Assisted Selection (Figure 2.8). Currently most research has been

geared towards Assisted Inspection, with algorithms designed to assist the operator, such as by annotating the image with a Region-Of-Interest (ROI) to prompt the operator of a potential security- or customs-related threat. Furthermore, image analysis can be used to improve image quality or compute material properties that can help operators identify threats. The goal of Assisted Selection is to use automated image analysis to inform the risk analysis used for cargo selection, but relies on the ability to scan all containers at high throughput rates. When high-throughput scanners are widely deployed, Assisted Selection has the potential to increase true positive and reduce false positive cargoes in the selected sample. In doing so, it should allow for human resources to be allocated more efficiently.

All commercial products are currently geared towards Assisted Inspection.

Automated image analysis has other advantages. Firstly, typical inspection algorithms give more consistent decision times than humans. For example, and in theory, a window-based inspection algorithm running on a dedicated computer system, will make a decision in a time

$$N_w \times t_w + t_a, \tag{2.1}$$

In a window-based algorithm, the image is split into a number of ROIs (windows) and each one is analysed individually, before the results are aggregated to determine the overall processing output for the image. The assumption that t_w is fixed, is true for most algorithms, however there are some CNN architectures that can exit early if the analysis is straightforward [58]. In this case, t_w is variable.

where N_w is the number of analysis windows in the image, t_w is the time taken to analyse each window, and t_a is the time taken to aggregate the window results to come to a decision on the whole container. And since the size of each container is known, the number of windows is known, and so is the analysis time for a given container. This is independent of the difficulty of the image content. Human speed, however, can vary due to a number of different factors, for example if they are tired, distracted, or the image is particularly complex, then image inspection can take longer. Consistent inspection times are beneficial for planning and budgeting resource allocation.

Secondly, it is easier to tune the Detection Rate (DR) and False Positive Rate (FPR) of an algorithm. This is important operationally, since the system can be tuned for different purposes. For example, if there was intelligence about an imminent terrorist threat using weapons, the system could be set to detect almost 100% of weapons but with a relatively high false alarm rate. Alternatively, a customs agency might be satisfied with detecting only 30% of cigarettes, so long as there were very few false positives. With automated inspection algorithms, the DR and FPR can be set using a tunable confidence threshold. With humans, it is much more difficult to control the DR and

FPR, but one might be able to achieve it by controlling the amount of time allowed for image inspection. Even so, this is difficult in practice.

Thirdly, computing power is easier to scale-up than human power. It just requires purchasing additional hardware and installation of relevant software. And it is easier to determine the extra resources available because of the more consistent analysis times in automated systems. Employing more humans can take time, and requires expensive training.

Finally, inspection algorithms do not suffer from a number of human factors and potential rogue external influences as outlined in Section 2.2.2.

2.7 SUMMARY

This chapter has identified the key security vulnerabilities in the cargo supply chain. Since high container throughput is of great economical importance, the thoroughness of container inspections has been restricted to only a small fraction of the container fleet. Security is weak, particularly when compared to aviation security, where all goods and people are inspected before transportation.

Historically, one aspect of the security protocol that has limited inspections was the use of NII to scan containers and their contents. However, recent advances in high-throughput scanner technologies will remove this limitation if such systems become widely adopted. The new limitation, is the sheer volume of images that need to be visually inspected, and the limited human resources available to do so. There are other limitations to relying solely on humans, including, that they are error prone, are inconsistent with themselves and with each other in terms of inspection speed and accuracy, and are vulnerable to external rogue influences.

In this thesis, methods of automated image analysis are proposed, which have the potential to reduce these limitations, whilst improving the throughput of containers.

3

LARGE-SCALE TRANSMISSION X-RAY IMAGING

This chapter introduces the principles of X-ray imaging, and how these are practically implemented at large-scale to image cargo containers and vehicles for the purpose security screening. X-ray vehicle and cargo scanners are different to typical X-ray devices used in medical imaging or for screening baggage at airports. For example, the required physical size of the apparatus introduces sources of noise not seen in other systems, and much higher X-ray energies are required to probe the container contents, meaning that the nature of the dominant X-ray interactions with matter are different to typical X-ray scanners from the medical or aviation security domains.

3.1 A BRIEF HISTORY

X-rays were accidentally discovered by several Physicists in the 1880s, as they often manifested as unexplainable fogginess or shadows on photographic plates placed near Crookes tubes [59]. However, it was not until 1895 that the first paper was published on X-rays by Röntgen. The paper excited the global Physics community, and many researchers had repeated Röntgen's experiments in a matter of weeks. Industry was quick to catch up, and before long several private X-ray imaging facilities were set-up and adverts for X-ray equipment began to appear in newspapers [53].

Röntgen referred to them as 'X-rays' to indicate that they were a new, unknown type of ray [60]. The name has become permanent.

As little as a year later, in Paris, Dr. T. Bordas suggested that X-rays could be used to interrogate the interiors of suspicious packages to check whether they were 'infernal machines' [53]. By 1897, customs houses in France had begun using X-rays to screen passenger belongings for contraband [60]. So surprisingly, the use of X-rays by customs officials began much earlier than many have reason to believe. Since then, there have been numerous technological advancements to reach today's X-ray instruments that are commonplace at ports around the world.

3.2 X-RAY PHYSICS

X-rays can, as can other types of electromagnetic radiation, be partly described in terms of waves and partly in terms of particles. This is known as the wave-particle duality. Each X-ray photon has an energy E determined by the Plank-Einstein relation

$$E = h\nu = \frac{hc}{\lambda}, \tag{3.1}$$

For example X-ray diffraction imaging exploits the wave-like nature of X-rays [61].

where h is Plank's constant, c is the speed of light, and ν and λ are the associated wave frequency and wavelength respectively. X-rays exhibit a number of particle- and wave-like phenomena which are exploited in a range of X-ray image devices.

Although high-energy 'X-rays' and 'γ-rays' can have similar energies, they are distinguished by the fact that X-rays arise from electron transitions, whereas γ-rays arise from nuclear transitions.

In large-scale transmission radiography for cargo and vehicle screening, energy cut-offs in the range 1-10 MeV are employed which is much higher than the energies of typical X-rays which range from 100 eV to 100 keV.

3.2.1 *X-ray generation*

The high X-ray energies are required in cargo screening to provide sufficient penetration. To achieve this, cargo scanners typically use a LINear ACcelerator (LINAC). The basic LINAC principle is as follows. Electrons are emitted by a cathode, and a high voltage source pushes the electrons through a hollow pipe vacuum chamber. Cylindrical electrodes of increasing length are placed within the vacuum chamber and given alternating charges. As electrons pass through they are repelled by the negatively charged electrode and then attracted towards positively charged electrode. In this way the electrons are accelerated until they collide with a tungsten target. In the collision they rapidly decelerate and create X-rays through the Bremsstrahlung process.

Bremsstrahlung is a German word, which roughly translates to 'braking radiation'.

The resulting X-ray energies fall into a continuous spectrum bounded by the voltage used to accelerate the electrons, which is known as the cut-off energy. An example of typical Bremsstrahlung spectra for 4 MeV and 6 MeV sources are shown in Figure 3.1.

A metal collimator is used to shape the beam into the required geometry. For cargo screening, a fan-beam geometry is often used to provide coverage of the whole cargo container or vehicle. A fan-beam has an approximately Gaussian-shaped cross section, and the width of this Gaussian, as determ-

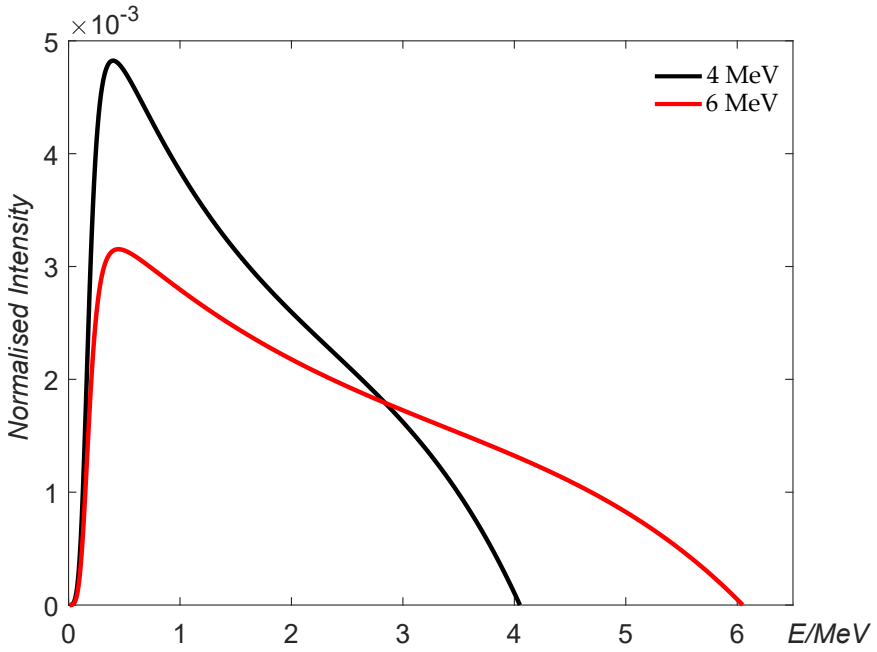


Figure 3.1: Simulated photon energy spectra for 4/6 MeV dual-energy X-ray cargo scanner; the two energies used in the scanners employed in this thesis. Generated using code supplied by James Ollier, Rapiscan Systems.

ined by the collimator, is chosen dependent on the configuration of the detectors. For systems with a linear detector array, it is important to make the width small so that the maximum number of photons hit the detectors, however, if too small and the detector is wobbling, the beam may miss the array completely. For systems with a planar detector array, the Gaussian width is increased to give approximately uniform signal across the array's width.

The effects of detector wobble are addressed in Chapter 5

Cargo scanners are operated in a pulsed mode, which means that a pulse of X-rays is emitted in a short period of time to sample the aligned point in the scene. For dual-energy, systems the pulses alternate between high and low energy, and the pulses are untangled to form both a high and low energy image. The pulse frequency is typically around 100 Hz, meaning that a typical dual-energy scan of a 40 ft container takes approximately one minute if a 1D linear detector array is employed. A high-speed scanner with a planar array can scan a 40 ft container in less than a second.

A linear array is used in Chapter 5 and a planar array in Chapters 7 and 8.

3.2.2 *Interactions with matter*

An X-ray beam passing through a material can be attenuated through a number of different particle interactions, including absorption and scatter. The amount of attenuation can be described by the Beer-Lambert law. Consider

X-rays of intensity I_z that are incident on a thin slab of homogeneous material with cross-sectional area A, thickness dz and concentration C. The number of atoms illuminated by X-rays is given by $CA\,dz$. The atoms have a cross section σ, which is an effective area that quantifies the intrinsic likelihood of an attenuating interaction with the beam. Therefore, the total effective area of the attenuating constituents in the material is given by $\sigma CA\,dz$. From this, the change of intensity dI_z across the material slab can be expressed as

$$dI_z = -I_z \frac{\sigma CA}{A} dz. \tag{3.2}$$

Integrating both sides yields the well-known Beer-Lambert law

$$I = I_0 e^{-\sigma C l} = I_0 e^{-\mu l}, \tag{3.3}$$

where l is the thickness of the material and μ is the attenuation co-efficient. Since the probability of interactions vary as a function of the energy of the incident X-ray photons, σ, and thus μ, have an energy dependence. Additionally, μ varies by the atomic number Z of the material. For inhomogeneous materials Z is dependent on z. So more generally, the Beer-Lambert law is expressed as

$$I(E) = I_0(E) \exp(-\int \mu(E, Z(z)) dz). \tag{3.4}$$

This is the basic principle that underpins X-ray imaging. Thick and dense (concentrated) materials attenuate more photons and appear darker in the image.

X-ray photons interact with matter via scattering and absorption by several mechanisms each with their own interaction cross section. Scattering alters the trajectory of the X-ray photon, whilst absorption destroys the original photon. The interactions most relevant to cargo X-ray imaging are photoelectric absorption (σ_{pe}), Rayleigh scatter (σ_{rs}), Compton scatter (σ_{cs}) and pair production (σ_{pp}). For each interaction, the cross section is dependent on the photon energy E and the characteristics of the material being imaged. The total cross section is the net effect of these different types of interaction, such that

There are many other possible interactions, but these have small probabilities (cross sections).

$$\begin{aligned} \sigma(E, Z) &= \sum_i \sigma_i(E, Z) && (3.5) \\ &\approx \sigma_{pe}(E, Z) + \sigma_{rs}(E, Z) + \sigma_{cs}(E, Z) + \sigma_{pp}(E, Z), && (3.6) \end{aligned}$$

where σ_i are the attenuation coefficients from the individual interactions.

A plot of the attenuation coefficient contributions for iron are plotted in Figure 3.2. At low energies the most dominant interaction in photoelectric absorption, and for higher energies pair production begins to dominate. Incoherent scatter dominates for intermediate energies. Although the Beer-Lambert law that underpins transmission X-ray image formation is the same for baggage and cargo, the interactions involved in the formation differ due to the difference in photon energies employed.

In the next sections, the interactions relevant to cargo imaging are described: photoelectric absorption; Rayleigh scattering; Compton scattering; and pair production.

Note that in Figure 3.2, the contribution of photoelectric absorption and Rayleigh scatter are several orders of magnitude less than Compton scatter and pair production at cargo cut-off energies.

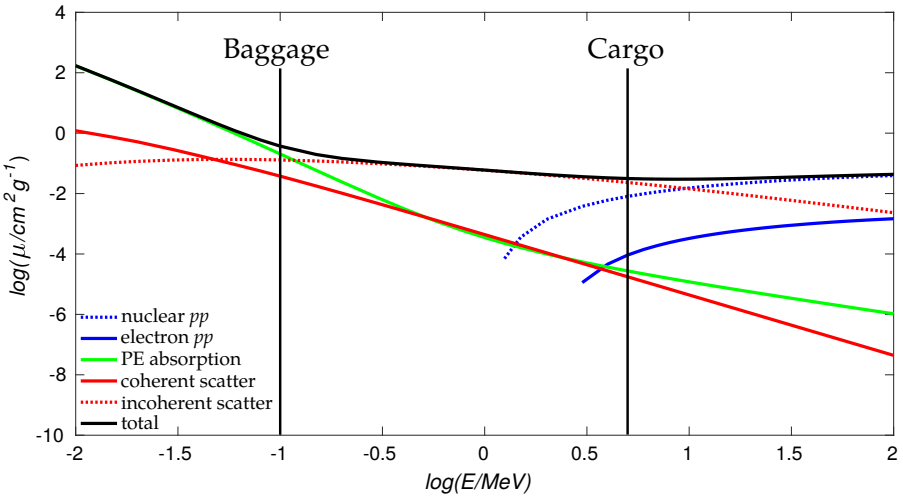


Figure 3.2: Attenuation coefficient (μ) contributions plotted for iron as a function of photon energy E. The vertical lines indicate typical photon cut-off energies used in baggage and cargo screening. Estimation of μ, which can be determined by the difference between images captured at two energies, is more difficult at cargo energies since the total μ gradient becomes small and large-scale commercial systems tend to suffer from severe noise. Data source: NIST [62].

3.2.2.1 *Photoelectric absorption*

The photoelectric effect is where the incident X-ray is absorbed by a bounded electron. The photon energy is imparted to the electron which results in an increase in its kinetic energy. At high energies, this extra energy is enough for the electron to escape the potential well of the nucleus, resulting in the ejection of the electron and ionisation of the atom. The electron will subsequently dissipate its energy in the material. If the electron was ejected from an inner orbital, it is often replaced by an electron from an outer orbital, emitting a photon as its energy is reduced.

Albert Einstein explained the effect in terms of quanta (photons) and it lead to his Nobel Prize in 1921.

The photoelectric cross section can be approximated as shown in Equation (3.7) [63].

$$\sigma_{pe} \sim \begin{cases} Z^4/E^3 & \text{low energy} \\ Z^5/E & \text{high energy,} \end{cases} \tag{3.7}$$

where Z is the atomic number. This equation implies that the probability of photoelectric interaction is higher at low energies than at high energies.

3.2.2.2 *Coherent scattering*

Coherent scatter, also known as Rayleigh scatter, occurs when the potential well is large relative to the energy of the photon. In other words, the electron is too tightly bound to the nucleus. In this case, the photon is absorbed and immediately re-emitted by the electron. No energy transfer occurs, but the photon can change direction after the process. Since the energy, and thus wavelength, of the photon does not change. This process is known as coherent scatter.

At X-ray energies the Rayleigh scatter cross section can be approximated as shown in Equation (3.8) [63].

$$\sigma_{rs} \propto (Z/E)^n, \text{ where } 2 < n < 2.5. \tag{3.8}$$

3.2.2.3 *Incoherent scatter*

Compton scatter was discovered by Arthur Compton in 1923. The cross section is dependent on the scatter solid angle Ω. This differential cross section $d\sigma_{cs}/d\Omega$ is given by the Klein-Nishina formula, which can be integrated to obtain the total cross section.

Incoherent scatter, also known as Compton scatter, is where the photon collides with a free or loosely-bound electron. In this case some photon energy is imparted to the electron, which is ejected if loosely bound. The total energy and momentum of the photon-electron system is conserved, however the photon often changes direction after the interaction. Inverse Compton scatter can also occur, where the photon emerges with increased energy.

In Compton scatter the change in photon wavelength is described by

$$\Delta\lambda = \frac{h}{m_e c}(1 - \cos\theta), \tag{3.9}$$

where m_e is the mass of an electron and θ is the scattering angle.

The total cross section for Compton scatter is expressed as shown in Equation (3.10) [63].

$$\sigma_{cs} \propto Z.\left\{\frac{1+\epsilon}{\epsilon^2}\left[\frac{2(1+\epsilon)}{1+2\epsilon}-\frac{\log(1+2\epsilon)}{\epsilon}\right]+\frac{\log(1+2\epsilon)}{2\epsilon}-\frac{1+3\epsilon}{(1+2\epsilon)^2}\right\}, \tag{3.10}$$

where $\epsilon = E/m_e c^2$.

3.2.2.4 *Pair production*

Pair production results in the complete attenuation of the incident photon. It occurs when a high-energy photon interacts with the strong electric field around the nucleus, spontaneously creating an electron and positron. In the process charge is conserved, and the photon energy is converted into the mass of the electron-positron pair. For these interactions to occur in the field of a nucleus, the photon energy E must be at least 1.022 MeV, which is twice the rest mass of the electron. For E>1.022 MeV the excess energy is shared, often unequally, between the electron and positron. Pair production is relevant to cargo scatter since the X-ray energy cut-off exceeds 1.022 MeV, however this is not the case for other X-ray imaging systems, such as those used to inspect baggage at airports or those used in medical imaging.

Note that pair production can create other lepton-anitlepton pairs, such as tau-muon. However, these have greater mass and require much higher energies than used in cargo screening.

E>2.044 *MeV in the field of an electron.*

The pair production cross section can be expressed as shown in Equation (3.11) [63].

$$\sigma_{pp}(E,Z) \propto \begin{cases} A\log(E) & \text{for low energy,} \\ A/X_0 \propto Z(Z+1)\log(287/\sqrt{Z}) & \text{for high energy,} \end{cases} \tag{3.11}$$

where X_0 is the radiation length of the material and A is the mass number.

Radiation length characterises how photons (and other radiation) interact in a material, and depends on the density and charge of the nucleus. It is defined as the mean path length required to reduce the energy of the particle by a factor of $1/e$. *This excitation process is known as luminescence*

3.2.3 *X-ray detection*

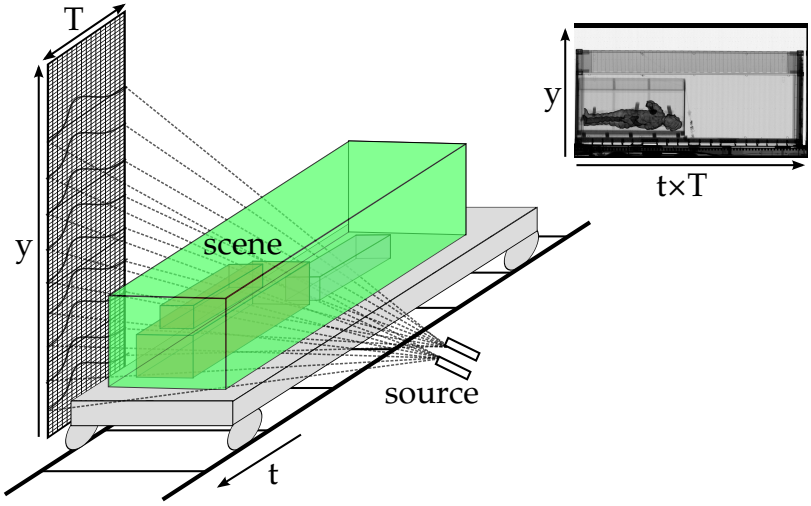
For X-ray detection, most commercial scanners use scintillator-photodiode detectors [64]. Scintillators are excited by the incident X-rays, and re-emit the absorbed energy as light, which is detected by the photodiode. For high energy X-rays, cadmium tungstate is often used [64]. The scintillating crystals can be several centimetres long.

Choosing the size of the detectors is a balancing game between the number of photons captured, and the spatial resolution in the image. By captur-

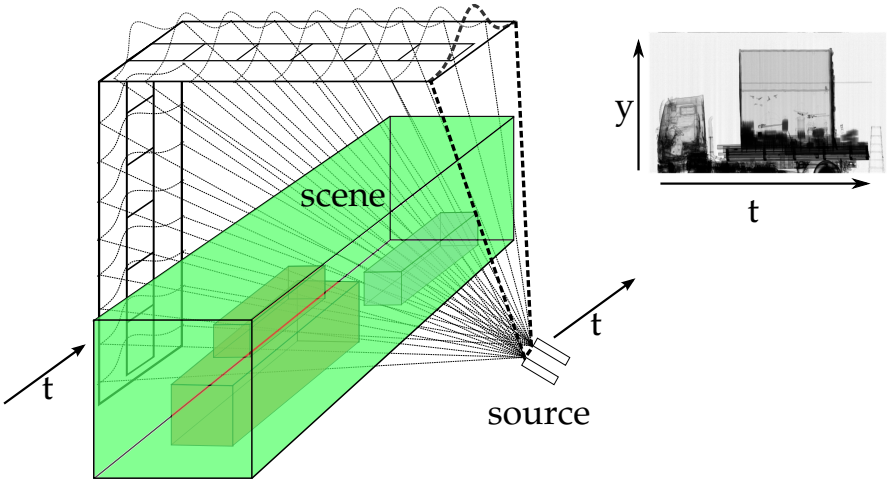
ing more photons, the Signal-to-Noise Ratio (SNR) can be improved but the spatial resolution is reduced, and vice versa. For cargo screening, the detectors are usually positioned in a linear or piece-wise linear (Γ-shaped) array (Figure 3.3b). The scintillating crystals are orientated so that each points towards the source. There are other systems which employ a planar detector array (Figure 3.3a), these are particularly useful for high-throughput scanning since rather than measuring a single image column at a time, it can measure T, so essentially the scanning can be done T times faster. However, such systems are more expensive since they require more detectors.

Note that we refer to it as 'Γ-shaped' rather than the 'L-shaped' typical of the literature.

For each X-ray pulse the scanned object can be moved on T pixels rather than one pixel.



(a) Portal mode (high speed)



(b) Traverse mode

Figure 3.3: Traverse and portal mode scanner architectures. The portal mode scanner has a planar detector array allowing for high-speed scanning, and the traverse mode scanner has a Γ-shape piece-wise linear array.

The detectors are connected to data acquisition electronics, which convert the photodiode readings into a photon count. This usually includes signal amplification, and analogue-to-digital conversion.

3.3 LARGE-SCALE TRANSMISSION RADIOGRAPHY

Since the LINAC produces a continuous spectra of X-rays, the image formation equation is given by integrating the Beer-Lambert law together with the energy spectrum

$$I = \int_0^{E_{max}} I_0(E) \exp(-\int \mu(E,z)dz)dE. \tag{3.12}$$

Often the response of a detector is dependent on the energy of the incident photons, and thus a factor is included to incorporate this, so that

$$I = \int_0^{E_{max}} I_0(E)D(E) \exp(-\int \mu(E,z)dz)dE, \tag{3.13}$$

where $D(E)$ is the response of the detector.

3.3.1 *Modes of deployment*

Large-scale transmission radiography scanners operate either in portal or traverse mode, and some systems are capable of both [65]. Illustrations of portal and traverse mode scanning architectures are given in Figure 3.3.

In portal mode the scanner is stationary and the container/vehicle moves between the source and imaging array at a controlled speed. In traverse mode the detector and source move either side of the stationary container/vehicle. Portal mode is most useful in high-throughput scenarios; vehicles can drive through the scanner arch without the driver having to exit the vehicle or a rail-scanner can scan multiple cargo containers carried by train at up to 60 km/h.

Some systems detect the driver's cabin and only activate the X-rays when they have passed through the scanning arch.

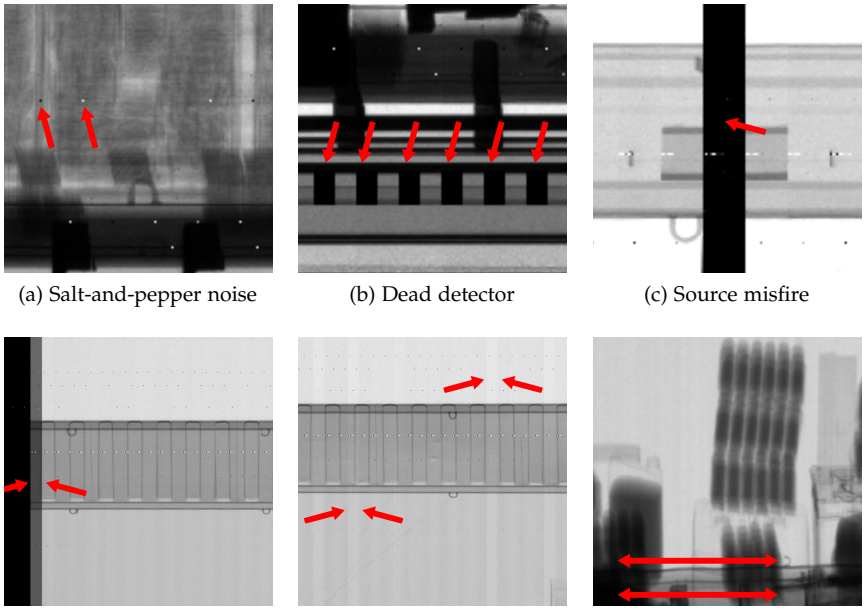
Traverse mode is useful in security scenarios where an unoccupied vehicle cannot be interfered with, such as if it is suspected to be a car- or truck-bomb, or if it needs to be covertly inspected. The traverse mode is also useful for scanning lines of stationary cargo containers at ports [65]. The traverse mode has advantages in some cases: (i) the scanned vehicle is unoccupied, so higher doses can be used, resulting in higher precision images; (ii) there

is greater control over scanning speed and detector-object distance resulting in less spatial warping of the captured image; and (iv) they have a compact scanning footprint [66].

Although some systems account for warping in image pre-processing.

3.3.2 *Noise, artefacts, and distortion*

There are several types of noise, artefacts, and geometric distortions that effect X-ray imagery and can reduce the ability of human operators or algorithms to detect threats. Their origins can be due to the physics of X-rays, or due to defects in the scanning equipment. Examples of different noise, artefact, and distortion are shown in Figure 3.4.



(a) Salt-and-pepper noise (b) Dead detector (c) Source misfire

(d) Source variation (e) Source variation (f) Geometric wobble

Figure 3.4: Example of typical image artefacts, noise and distortion that appears in large-scale transmission radiography. Red arrows indicate effects.

The Binomial or Poisson selection of a Poisson process is also Poisson.

POISSON NOISE The physics of X-ray imaging involves several random processes, and when combined they result in Poisson noise in the image pixels. The random variation can include: the number of photons emitted by the source; the number of photons that pass, unattenuated, through the scene; and the number of photons captured in the scintillating crystal. Since these random processes are Poisson or Binomial, the overall process results in Poisson distributed noise. One of the consequences of Poisson noise is that the variance in photon counts is equal to the expected photon count.

This means that for a given homogeneous object in the scene, the SNR varies as the square-root of the mean pixel count.

SALT-AND-PEPPER NOISE X-ray images in digital systems can also suffer from salt-and-pepper noise. This can be due to pixels that go temporarily offline, errors in analogue-to-digital conversion, or data transmission. It presents itself as sparsely occurring white and black pixels. Some systems tend to suffer from more salt-and-pepper noise than others. Figure 3.4a shows an example of salt-and-pepper noise.

In this thesis, the railscanner system employed in Chapter 7 suffers from more severe salt-and-pepper noise than the gantry system employed in Chapter 5.

WOBBLE Detector wobble only affects systems operating in traverse mode, where the detector can wobble as the system moves due to mechanical instability. It can manifest itself as noise, artefacts, and geometric distortions. For example in Gantry systems, where the scanner moves along rails, the pattern of wobble is generally consistent, affecting each image in a similar way, and thus wobble creates image artefacts. In these systems, it is generally caused by the rails being bent or buckled, or due to irregularities in the driving mechanism. In truck-mounted traverse-mode systems, wobble is unpredictable due to heavy gusts of wind or variations in the ground topology when scanning at different locations. Thus, wobble is a random process which is different for each image, and wobble presents as image noise. In cases of extreme wobble, there can be geometric distortions in the image. This is most obvious where straight lines appear to be bent, when they should not. Wobble artefacts and noise are addressed in Chapter 5, and an example of geometric distortion due to wobble can be seen in Figure 3.4f.

SOURCE VARIATION AND MISFIRE Since each LINAC pulse forms a column (or number of columns for planar arrays) of the scene, source variation can manifest as vertical stripes in the image. This is observed in most cargo systems. Examples of source variation can be seen in Figures 3.4d and 3.4e. Source misfire occurs when the source does not fire for one or more pulses during image acquisition, in this work this is particularly noticeable for the railscanner used in Chapter 7, since it is a planar array; one missed pulse leads to noticeable missing image content such as in Figure 3.4c.

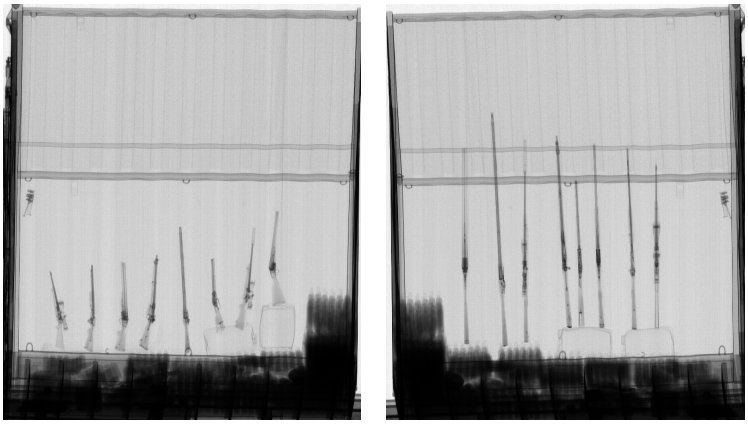
DETECTOR ARTEFACTS Detector artefacts can lead to row artefacts in the image. This can be due to the variable distance of the detectors from the

source, variable detector orientation, variable detector displacement from the array mid-line, or if groups of (or single) detector pixels are dead (Figure 3.4b). Variable distance, orientation and displacements are addressed in Chapter 5.

SCATTER Scatter through the Compton process leads to a blurring effect in images, particularly noticeable at the edges of objects. Some systems attempt to correct for this by using an anti-scatter grid to block photons moving along scattered trajectories.

3.3.3 *Image properties*

Cargo X-ray imagery have some unique properties. Due to the large-scale of the volumes that are imaged and the divergent X-ray fan-beam, a magnification effect occurs. The magnification depends on the distance of an object from the source. If it is close to the source it intercepts more lines of photon flux, since they are bunched closer together near the source. If it is further away, it intercepts fewer lines of flux as they have a longer distance over which to diverge. This results in objects placed close to the source appearing taller than those placed further away, as shown in Figure 3.5. Thick objects (e.g. the container itself) appear skewed in the image due to the relative magnification between the front and back of the object. Similarly long, straight, thin objects can appear bent if oriented so that one end is close to the source and the other is far way.



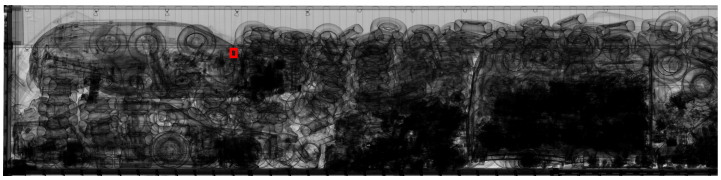
(a) Far from source (b) Close to source

Figure 3.5: Example of the magnification effect in X-ray cargo imagery. Weapons placed furthest from the source (a) appear much shorter than when placed close to the source (b).

The projective nature of X-ray imagery is, in theory, beneficial over optical imagery, since objects appear translucent. This means that if an object occludes another, the information about both objects is encoded in the image unless one is completely radiopaque. However, this can make images very confusing, particularly if they contain lots of complicated overlapping objects. This can limit the human operator's capability to quickly locate and identify potential threats.

Typically, cargo X-ray scanners capture 16-bit images with high dynamic range and effective spatial resolutions of a few mm/pixel [64]. This means that an uncompressed single-energy, single-view, X-ray image of a 40 ft container consists of approximately 2600×850 pixels and occupies around 5 MB of disk space.

(a) Airport baggage



(b) Cargo container

Figure 3.6: Comparison of a typical dual-view X-ray image of baggage and a typical single-view X-ray image of cargo. The red boxes indicate the approximate size of a handgun in the image. In baggage, the handgun occupies a large portion of the image, and there is little distracting clutter, so that there is a strong visual signature. In cargo, there is typically a lot more clutter and the handgun occupies a very small portion of the image, which makes handgun detection difficult for both humans and algorithms.

3.4 COMPARISON WITH BAGGAGE

Cargo images pose a difficult visual search task for the human operator, and they are much more difficult to analyse than other types of border security imagery such as airport baggage or parcels. This is because cargo scanners have to operate at a much larger scale. For example, a 40 ft General Purpose cargo container has a volume of 67.6 m^3 [67] and is made out of steel,

Determined based on British Airways cabin bag size allowance of 56 cm × 45 cm × 25 cm.

whereas hand luggage volume is typically $0.063\,\mathrm{m}^3$ and usually made out of fabric or plastics. The physical scale of cargo scanners makes it difficult to efficiently perform 3D Computed Tomography (CT) [51] but some multi-view systems do exist. Moreover, for cargo it is more difficult to extract material composition information due to the higher energies required for sufficient penetration to obtain good image contrast (Figure 3.2). Cargo images are also far more cluttered, whilst small threats, such as firearms, have a very small visual signature. A comparison between dual-view baggage and single-view cargo imagery is shown in Figure 3.6.

3.5 SUMMARY

In this chapter, the main principles of X-ray physics have been introduced, and their roles in image formation, noise, artefacts and distortions, were described. Knowledge of these physical processes plays an important role in the technical chapters of this thesis. In particular, understanding image formation allows us to construct a model of image formation in the presence of a wobbling detector in Chapter 5, which is used in the correction of wobble artefacts in cargo imagery. Further, the understanding of X-ray images and their natural variations leads to the development of a Threat Image Projection (TIP) framework (Chapter 6) and image pre-processing methods that are fundamental to Chapter 7 and Chapter 8.

4

LITERATURE REVIEW

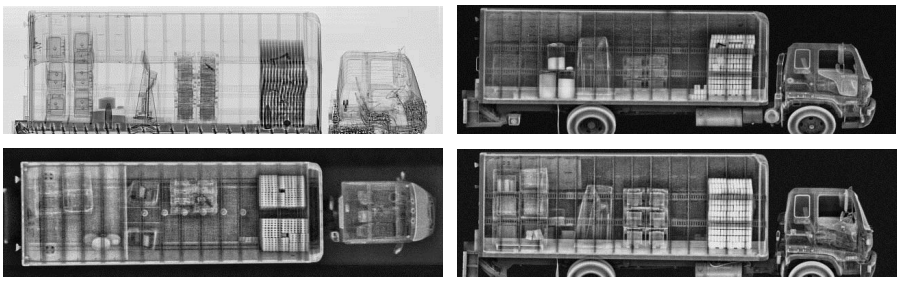
Here, we explore the prior literature on automated image analysis for cargo. The literature can be separated into the themes of *image pre-processing* (Section 4.2) and *image understanding* (Section 4.3). *Image pre-processing* is a broad category including any operation made to an image in order to help *image understanding* by either humans or algorithms. *Image pre-processing* includes: image manipulation; image correction and denoising; material discrimination and segmentation; and Threat Image Projection (TIP). *Image understanding* methods make decisions based on the image contents. Current methods for this can be split into Automated Threat Detection (ATD) and Automated Contents Verification (ACV). Section 4.4 summarises the state of the field and highlights how this thesis contributes towards moving it forward.

In some cases, the literature directly relating to cargo imagery is scarce. This is due largely to commercial and security protection, and the difficulty for academics to obtain access to commercial scanning hardware. Moreover, the majority of funding goes towards aviation security, where search tasks are more tractable, and there is a more obvious and immediate threat from terrorism. In cases where cargo research is sparse, other domains such as baggage can provide insight, since many of the findings there may be transferable to the cargo domain.

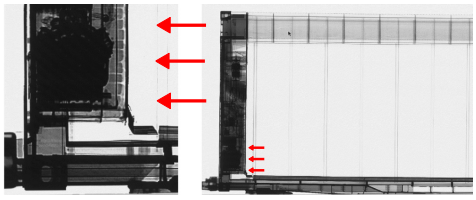
4.1 EXAMPLE X-RAY IMAGERY

Automated image analysis has been applied mostly to transmission X-ray imagery, however some researchers have conducted research on joint transmission and backscatter systems [68]. Examples of X-ray imagery from the literature are presented in Figure 4.1 along with the imagery used in this thesis.

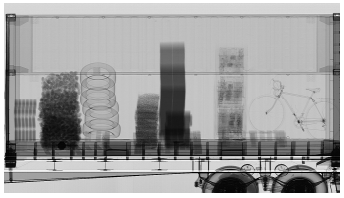
Backscatter systems differ from transmission system in that they measure X-rays reflected by the scene and back towards the source. This means that the detector and source are located in close proximity. The reflections oc-



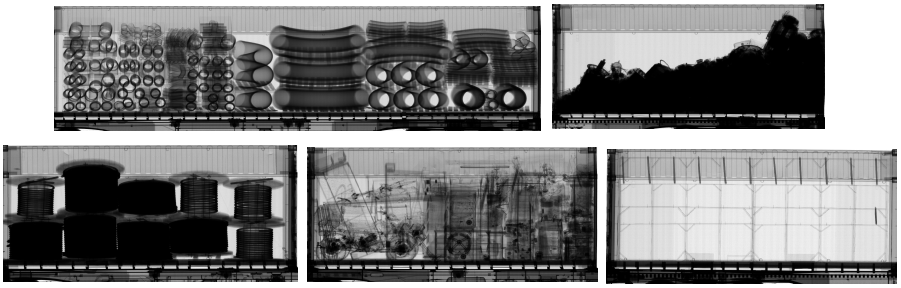
(a) Zheng and Elmaghraby [68]



(b) Vogel [69]



(c) Chalmers [70]



(d) Rapiscan Systems

Figure 4.1: Examples of cargo X-ray images, including: (a) triple-view backscatter and single transmission images from an AS&E OmniView® Gantry; (b) example of concealed cigarettes (indicated by red arrows) in a transmission X-ray image of refrigerated container; (c) transmission image of 20 ft container carried by lorry; and (d) examples of (pre-processed) transmission images from the SoC captured by a Rapiscan Eagle® R60 railscanner, and used for the work in Chapters 7 and 8.

See discussion on Compton scatter in Chapter 3.

cur primarily through Compton scatter. Backscatter images, do not provide high penetration, but are more suitable for detecting low density contraband, such as narcotics or cigarettes. Due to the low penetration, the cargo container often has to be imaged from multiple views, requiring extra detectors and sources, thus making them more expensive than transmission systems.

This review focuses primarily on single-view transmission X-ray imagery. Such systems provide much greater penetration and high contrast, making them more useful for detecting concealed items.

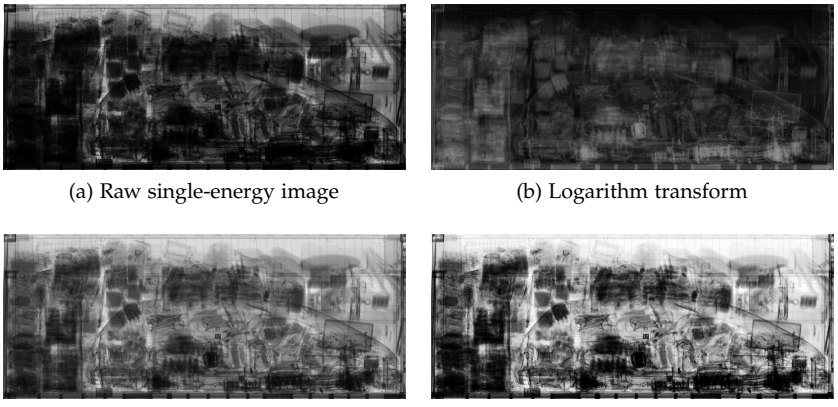
4.2 IMAGE PRE-PROCESSING

In this review, *image pre-processing* is defined as any process which is performed before, and in order to improve the performance of, *image understanding* by humans or automated systems. Four topics from the literature have been identified: image manipulation; image quality improvement; material discrimination and segmentation; and TIP.

4.2.1 *Image manipulation*

Image manipulation is used to improve the performance of human operators and potentially automated *image understanding* algorithms. Most work has been on studying the threat detection performance of human operators under different image manipulation functions implemented in commercial image viewing software. Manipulations include pseudo-colouring, edge enhancement, and intensity transforms such as Adaptive Histogram Equalisation (AHE), logarithm and square-root (Figure 4.2) [71].

Note that pseudo-colour, in this case, is not based on material properties. Material discrimination is discussed in Section 4.2.4.



(a) Raw single-energy image

(b) Logarithm transform

(c) Square-root transform

(d) Histogram equalisation

Figure 4.2: Examples of different image manipulations for an X-ray image of a car with partial occlusion. There is still some debate over which manipulation facilitates best human or algorithmic *image understanding*.

For cargo screening, Michel et al. [72] have found that image pseudo-colouring does not lead to improved performance over the raw greyscale, for humans searching for narcotics, weapons, Improvised Explosive Devices (IEDs) and other explosives. Similar results have been found by Klock [73], who tested human performance at detecting concealed IEDs, guns, knives and other prohibited items in baggage. Evaluated manipulations included pseudo-colour, intensity inversion, inorganic or organic material stripping,

Details of Crystal Clear™ are difficult to find, but the function 'optimises image contrast and resolution to bring out picture details' according to a public verbal communication by Andreas Kotowski (Rapiscan Systems CTO) in 2001.

and a commercial Crystal Clear™ image enhancement. They found that the raw grey-scale image and the Crystal Clear™ enhancement enabled best human performance.

Chen [64] reasons that although most X-ray cargo images are captured and encoded in 16 bits, typical grey-scale displays only use 8 bits and so useful information is lost, but with pseudo-coloured images, there are 8 bits available for each colour channel, thus potentially preserving the information. However, he argues that the effectiveness of pseudo-colour is in fact limited by the ability of humans to detect subtle colour differences. The author also claims that edge enhancement techniques do not work well for cargo due to the complexity of objects and high pixel noise. Chen [64] qualitatively evaluates linear, logarithm and AHE image transforms. He argues that the log transform can be beneficial as it makes image brightness approximately proportional to object thickness, but thin items are sometimes lost. Chen [64] states that the square-root transform can be beneficial since the Signal-to-Noise Ratio (SNR) is proportional to the square-root of pixel intensity, thus it is an equal-noise display method. Finally, the author observes that qualitatively AHE is the best method, but that information on object thickness is lost.

To best knowledge, there have been no specific studies on the effect of image manipulations on automated *image understanding* in cargo. However, other researchers have applied different image manipulations before applying *image understanding* algorithms. These include: Gaussian blurring [6]; rudimentary segmentation algorithms to extract different image regions [6]; Weiner filtering and wavelet shrinkage denoising [74]; and intensity inversion followed by z-score normalisation and Retinex filtering [68]. Retinex filtering is an image enhancement method that aims to mimic human visual perception, and the name 'retinex' is derived from the words 'retina' and 'cortex'. The method adjusts local contrast whilst retaining global tonal quality.

4.2.2 *Wobble and image quality improvement*

Image quality improvement can include denoising methods to ameliorate Poisson or salt-and-pepper noise, and methods to correct image errors that arise during image acquisition. A major component of this thesis is dedicated to the measurement and correction of detector wobble. There have been

no previous publications addressing wobble in large-scale transmission radiography. However, wobble leads to artefacts in a range of imaging devices, including C-arm Computed Tomography (CT) micro-CT. The most relevant works and how they relate to the wobble effect that this thesis attempts to measure and correct in Chapter 5, are discussed here.

C-arm CT systems suffer from wobble as the gantry rotates. This means that individual projections are translated relative to those captured by the ideal wobble-free device. Researchers note that the wobble of the C-arm gantry is often repeatable over periods of up to two years and so wobble artefacts can be corrected by a one-off system calibration [75, 76]. This is similar to some large-scale transmission systems, particularly those that are in fixed deployment and the gantry moves along rails, where the wobble effect tends to be systematic. However, in truck-mounted mobile systems, wobble is much more unpredictable due to variable scan speed and due to variations in the surface topology that the truck traverses and the dynamics of the truck suspension. Moreover, in C-arm CT, the wobble artefacts tend to lead to a blurring effect in the reconstructed image due to the misalignment of individual projections, whereas in large-scale transmission systems, wobble mostly leads to image intensity variations as the fan-beam comes in and out of alignment with the detectors.

Silver et al. [75] propose a method for determining and correcting wobble in C-arm CT. The authors assume that the wobbling motion of the C-arm is identical for each image capture process, and so calibrate wobble correction based on a phantom image. The phantom consists of a helical structure of tungsten carbide spheres (pellets). The calibration computes wobble coefficients that are used directly in image reconstruction to obtain an image free from wobble artefacts. The wobble coefficients are determined by fitting a mapping from physical space to projection space using least-squares. Fahrig and Holdsworth [76] also adopt a calibration approach to determine projection translations. They use a bicubic spline interpolation to determine translated projections. Since the calibration process determines translations at discrete gantry angles, they linearly interpolate between them to obtain estimates for different projection angles if required.

Wobble is also observed in micro-CT systems, but the wobble manifests in the rotation table since the detector and source are kept stationary [77]. In this case, wobble again leads to a blurring effect in the image, quite different to the effect observed in large-scale transmission systems. Authors have

investigated image-based, calibration-based, and online methods to correct for wobble.

Sasov et al. [77] investigate and evaluate both an image-based method and a calibration-based method. The image-based method is an iterative compensation scheme, which first does an initial reconstruction using filtered back-projection, yielding a blurry wobble affected reconstruction. Estimates for projection translations, to compensate for wobble, are determined by comparing the original projections with corresponding forward-projected image estimates. The comparison is done either by cross-correlation or least-squares. Under these translations a new reconstruction is made and the process is iterated until the reconstructed image is satisfactory. The calibration-based method, measures wobble in a short reference scan directly before or after image capture to determine the compensatory translations of individual projections. They measure the position of the focal spot, relative to a metal pin placed in the scene, by fixing a fine metal mesh to the X-ray source. The authors claim that the second method is more suitable for slow drifts (wobble) and that the approach is faster and less computationally demanding than the iterative based method. However, the image-based method has the advantage of working purely on measured image values.

Zhao et al. [78] propose an online method, which uses capacitive distance sensors to measure the wobble of the rotation table in micro-CT. The measurements are used to translate individual projections to compensate for the displacement of the rotation table due to wobble. The authors report that the methods improve image quality by 53.1% and 65.5% when calibrating projections in the horizontal and vertical directions, respectively.

To best knowledge, there have been no published comparison studies on different cargo image denoising techniques. However, in baggage, Mouton et al. [79] perform a comparative study on a number of denoising techniques applied to low quality baggage imagery. The techniques included:

(i) Total Variation (TV) denoising – an approach to image denoising that uses total variation regularisation. The method smooths noise in the image, whilst offering better edge preservation than simple techniques such as median filtering.

(ii) Translation Invariant Wavelet Shrinkage (TIWS) – an approach to image denoising that transforms the image into the wavelet domain, and then applies a threshold on the wavelet coefficients. Translation invariant

wavelets are used to help overcome visual artefacts that are induced when using translation variant wavelets.

(iii) Non-Local Means (NLM) filtering – As opposed to local mean filtering, a non-local mean filtering takes a mean of all pixels in an image weighted by how similar each pixel is to the local pixel. In this way, NLM, is able to preserve more detail than a local mean filter.

The authors assess performance by running a Scale-Invariant Feature Transform (SIFT) point detector across image before and after the denoising. They identify object feature points (located on an object of interest) and noise (not on the object) within the image. Performance is then measured by taking the number of object feature points as a fraction of the total number of feature points, assuming that an increasing ratio is indicative of improved performance using a SIFT-based detection algorithm. They find that all methods offer improved performance, over using just the raw image, with TIWS performing best. However, it is unclear whether these results would generalise to algorithms that are not based on SIFT.

4.2.3 *Threat Image Projection and data augmentation*

TIP is a technique first developed for baggage [80]. Most TIP methods insert a Fictional Threat Image (FTI) from a database into an existing benign image. This can be used in Computer-Based Training (CBT) of operators, to assess performance [81], or improve it by increasing an operator's exposure to rare threat scenarios [82]. Similarly, in cargo, researchers are exploring how TIP imagery can be used to increase the competency of operators, however, so far they have relied on using screening experts to manually merge *threat* and *benign* images to form the TIP image using X-ray image merging software [72].

In CT baggage, TIP is complicated by the 3D nature of the images. TIP algorithms typically search for realistic placement volumes (voids) [83, 84] so that the projected threat does not intersect other objects which would create an unrealistic visual cue for operators. Researchers have defined metrics for View Difficulty, Superposition, and Bag Complexity [85–87]. Such metrics could be used for adaptive CBT algorithms, where the difficulty of a given search task can be controlled. For example, if an operator is poor at finding threat items in certain contexts, such as complicated clutter, the algorithm can present more of these examples to improve performance under those

contexts. Other researchers have noted that TIP imagery appears unrealistic unless realistic noise and artefacts, that match those of the other objects in the baggage image, are synthesised. For example, Megherbi et al. [83, 88] generate realistic metal artefacts in CT baggage, to ensure that artefacts in the threat are consistent with those in the rest of the baggage, and are not a visual cue for operators. Similar ideas may be useful in cargo TIP, for example ensuring that magnification and pixel noise is consistent between the threat and the rest of the image.

In cargo, some authors have proposed image synthesis methods for other purposes, but which could be used for TIP. White et al. [89] introduce a method for generating synthetic γ-ray cargo images, and use it as surrogate data for testing the effectiveness of different scanning systems, where it is impractical to collect large amounts of empirical data. The authors derive an empirical model of the imaging system response from real images of well-characterised objects. They claim to incorporate system properties such as sensitivity, spatial resolution, contrast and noise. To synthesise a threat image, the authors simulate photon transmission using a commercial ray-tracing package, and then apply smoothing and Gaussian noise consistent with their empirical measurements. The ray-tracing software allows for simulation of complex-object models, such as those developed in Computer Aided Design (CAD). After simulating the photon transport and detector-response model the synthetic threat images are injected into real images. They perform this injection by pixel-wise multiplication of the synthetic threat image with the real image. It is possible, that synthesising threat images from 3D threat models could prove useful, particularly for adding emerging threats to TIP libraries, for example CAD models of 3D-printed firearms.

TIP from 3D CAD models is not attempted in this thesis, but is a potential avenue to explore in future research.

In the Machine Learning (ML) community, training data augmentation is used to improve the performance of ML-based algorithms. Data augmentation reduces overfitting by using label-preserving transformations to artificially enlarge the dataset [90]. Transformations must be realistic for the given imaging system, in order to make algorithms robust to natural variation. In visible spectrum imagery, examples of transformations include random crops (translation invariance), random flips (reflection invariance), and random addition of lighting (invariance to illuminance) [91]. In cargo X-ray imagery transformations could include, for example, variations in dose, perspective, material composition, and object orientation. Data augmenta-

tion is particularly useful in representation-learning methods, such as deep Convolutional Neural Networks (CNNs), which are prone to overfitting if datasets are limited in size and variety.

So far there is no evidence that TIP has been used as a form of data augmentation for training and evaluating ML-based *image understanding* algorithms.

4.2.4 *Material discrimination and segmentation*

Material discrimination aims to identify the type of material at each pixel in the image. There is some crossover with *image understanding*, but it is included as an *image pre-processing* method herein, since *image understanding* methods using features derived from the material information might be helpful in improving performance. This has been the case in multi-view X-ray baggage, where material information is more complete [92]. The best performing methods of material discrimination also use segmentation approaches to refine the discrimination [5]. This is because high levels of noise make pixel-wise discrimination inaccurate, unless the material discrimination is averaged over segmented images regionsd.

The interactions of X-rays with a material varies depending on the type of material and the type of radiation (Section 3.2.2). By studying the types of interactions occurring, it is possible to identify the type of material by some characteristic such as its atomic number. To do this, it is required that measurements at multiple energies are made on the material either by illuminating it with multiple radiation sources [5], or by using a continuous spectrum of radiation energies and a detector that can resolve the difference in the energy spectrum after interaction [93].

Often the high-throughput requirements of commercial systems prevents more than two energies being used, permits at most a few views being acquired, and requires short exposures leading to substantial image noise. The combined effect of these restrictions makes the signal insufficient for discrimination of individual atomic elements [74]. Instead researchers attempt to discriminate between groups of materials such as organics, light metals and heavy metals [5, 94]. Alternatively some researchers attempt to only identify high-Z materials [74, 95, 96] as they can indicate the smuggling of radioactive materials or their shielding. Even in these simple cases, researchers have found it difficult to accurately discriminate materials from

raw measurements on a pixel-wise basis, finding that it is necessary to incorporate spatial information into discrimination [5]. Thus, researchers have applied a number of image segmentation approaches to aid with discrimination.

The majority of the cargo material discrimination literature uses dual-energy X-ray systems and are based on the α-curve [97, 98], R-curve [5], or H-L curve [99] methods. The physics behind these methods are described in more detail in Chapter 8 There has been little influence in cargo work from the baggage or medical domains, due to the much higher energy regime (see Figure 3.2). Take, as an example, the seminal work in CT by Alvarez and Macovski [100], which expands the attenuation coefficient as a set of intuitive basis functions. This approach works in the CT energy regime where the photoelectric interaction, which depends strongly on atomic number, is dominant. However, it is subservient to pair production and scatter in the cargo energy regime.

The α-curve [97, 98], R-curve [5], and H-L curve [99] methods attempt to estimate the effective atomic number (Z) grouping (i.e. organic, light metals, heavy metals) by combining high and low energy radioscopic transparencies T to form a value that can be mapped to effective Z grouping using a lookup table. Authors tend to define the transparency T by normalising the image by the total number of photons (integrated over the range of energies E) emitted by the source and the detector response $D(E)$.

The R-curve method is motivated by capturing transparencies T_1 and T_2 at energies E_1 and E_2, and taking the ratio of their logs

$$R(E_1, E_2, Z_0) = \frac{\log(T_1)}{\log(T_2)} = \frac{\mu(E_1, Z_0)}{\mu(E_2, Z_0)} \frac{D(E_1)}{D(E_2)}. \tag{4.1}$$

For the monochromatic and single material case, the R-ratio is unique to the material atomic number Z_0 and so materials can be discriminated, at least in theory. This method is well-suited to γ-ray imaging where the photons are emitted with quantised energies. However, in cargo, the X-ray source is not monochromatic and has a continuous Bremsstrahlung distribution. In this case R varies as a function of the material mass thickness. Nevertheless, one can attempt to recover the effective atomic number grouping at a pixel by experimentally measuring the R-ratio as a function of mass thickness to create a lookup table. There are difficulties at low mass thickness where the R-ratio versus mass thickness curves for different materials overlap.

The α-curve method uses the quantities:

$$\alpha_1 = -\log T_1, \tag{4.2}$$

$$\alpha_2 = -\log T_1 + \log T_2. \tag{4.3}$$

Again a lookup table is determined through experimentation.

Note the change in notation from the original publications where $\alpha_1 - \alpha_2$ was used in place of α_2 [97, 98].

Finally, the H-L curve method simply uses a lookup table of the high (H) and low (L) energy transparencies T_1 (H) and T_2 (L).

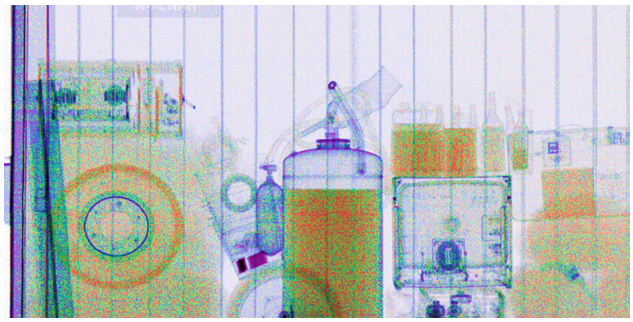
The seminal work for dual-energy material discrimination for cargo was by Ogorodnikov and Petrunin [5] and Ogorodnikov et al. [94]. The authors introduce the R-curve method and attempt to classify materials into four groups: organics (hydrocarbon, Z~5.3); organics/inorganics (aluminium, Z ~ 13); inorganics (iron, Z~26); and heavy substances (lead, Z~82). They use a prototype inspection system, with a 4/8 MeV cut-off Bremsstrahlung beam and a lead beam filter. They identify that the R-ratio crossover of iron and lead can be translocated by use of the filter, thus allowing improved discrimination for small mass thickness [5]. The authors first study the error, when discriminating iron from hydrocarbon, as a function of mass thickness, and find discrimination is optimal at 40-60 g/cm^2. They reason that discrimination error increases for lower mass thickness because there is not sufficient contrast between low and high energy images, and for larger mass thickness due to the decreasing SNR. The authors note that, when discriminating between all four groups, material recognition is unreliable. In particular, the water-aluminium discrimination error reaches 40% even at the optimal mass thickness. As a remedy, they incorporate spatial information using a spatial clustering algorithm. All pixels within a given cluster are labelled as a single material based on a lookup of the R-ratio table of the cluster mean values. Coloured material discrimination images, with and without incorporation of the spatial information, are shown in Figure 4.3. Qualitatively, it is evident that the use of spatial information greatly improves image quality.

A few years later, Zhang et al. [99] proposed the H-L curve method. They introduced a *material intrinsic difference* measure, defined as

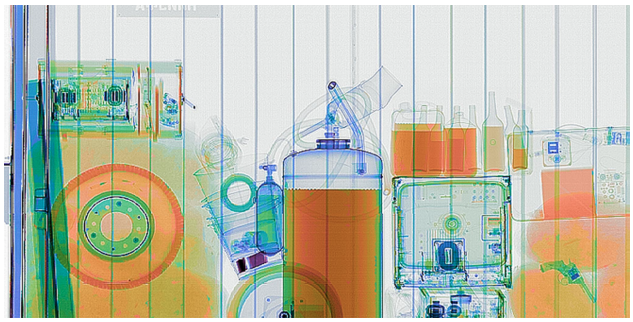
This is similar to a measure introduced by Ogorodnikov and Petrunin [5] and Ogorodnikov et al. [94] but for the R-ratio.

$$diff = \frac{H_2 - H_1}{(H_2 + H_1)/2} \times 100\%, \tag{4.4}$$

where H_1 and H_2 are the high energy measurements for two different materials. They use *diff* for evaluating the material discrimination abilities of a dual-energy system. The authors argue that if the image noise has mag-



(a) Material discrimination on individual pixels



(b) Material discrimination on spatial regions

Figure 4.3: Source: Ogorodnikov and Petrunin [5]. Material discrimination based on spatial regions provides improved image quality. This is particularly noticeable for the pistol, whose details are revealed in (b).

nitude less than *diff*, then materials can be accurately discriminated. They give a table of results showing the measured *diff* for different adjacent-Z materials and find *diff* to be a decreasing function of L. The authors do not show any results of applying the H-L curve method to whole images.

Since these initial works, other researchers have largely focussed on high-Z detection, claiming that multi-group material discrimination is infeasible for commercial systems. For example, Fu et al. [74] claim that identifying the effective Z of the scanned objects is not practical because it requires high precision measurements and the noise in commercial systems is too large. Most have focussed on the detection and segmentation of suspicious or high-Z objects.

Fu et al. [101] attempt to segment suspicious, shielded objects. They introduce a *hybrid clustering* approach which does not require a prior on the number of clusters or the size of clusters, but a prior on the *step level*, which determines the number of quantisation levels in the clustered image given the maximum image value. Hybrid clustering performs clustering followed by region growing. For clustering, each pixel is first compared to the mean of its neighbourhood, if the pixel is close to the mean then its value is assigned as the quantisation of that mean. If it is not close, then they split the

neighbourhood into quadrants, compute the means, and set the pixel value to the nearest quadrant's quantised mean. They claim that this is faster than recursive K-means clustering and the Leader clustering used by Ogorodnikov and Petrunin [5, 94]. After clustering they do region merging, using the highest intensity region as the seed. To segment shielded objects, the authors iterate through the different quantisation levels, binarise the image by quantisation level, and then region fill based on gradients. If the intensity of a filled region is greater than the surrounding, then it is regarded as a shielded object. The method is tested on a cargo image with various amounts of shielded lead and tin. No quantitative measure of the performance is given, but the method appears to work well on the single test image presented.

In a separate paper, Fu et al. [95] attempt to improve detection and reduce false alarms for high-Z detection. They apply their hybrid clustering approach described in [101]. After identifying regions that are shielded by low-Z materials, they attempt to separate the shielded object from the background by subtracting the *shielding* attenuation from the *shielded* attenuation. The authors claim that the approach yields improved high-Z detection. In a further paper [74], they identify several sources of error, namely; the *edge effect* at object edges due to scatter, misalignment, analogue-to-digital conversion, and Poisson noise. They propose a wavelet shrinkage denoising approach, which reduces false negatives and false positives, but no quantitative measure of performance is determined. The authors state that similar results can be achieved by use of a Weiner filter, but that it needs to be combined with morphological filtering.

Chen et al. [96] also focus on detecting high-Z material, using a 6/9 MeV commercial system. No substantial details of the methods are given, although they state that the high-Z signature is generated using 'dual-energy information processing, machine vision and topology analysis, and background object stripping' [96]. They show an example of lead detection against a piece-wise varying background density, but no quantitative measure of performance is given.

In a more recent paper, Ogorodnikov et al. [102] refer to their original work [5] and echo the sentiments of other researchers; that their previous approach to material discrimination is labile, unstable and not repeatable in practical implementation. In this paper, although their algorithmic methods are not detailed in full, the authors attempt 3-group (organics, mineral/-light metals, metals) material discrimination, but this time with a 3.5/6 MeV

Bremsstrahlung beam. Additionally, they attempt to calculate the mass of the object under inspection. They claim a mass precision of $<10\%$ and effective atomic number precision of ± 1 for materials in the optimal mass thickness range.

Recently, Li et al. [98] have proposed a solution to improve material recognition in cargo, when two materials overlap in an image. The method requires prior information about one of the overlapping materials, which the authors argue is available in a commercial setting from the shipping manifest, or if trying to separate container and contents an assumption can be made about the container material. Their algorithm firstly performs a pre-classification based on the α-curve method, they then determine if a region is more likely composed of a pure material or two overlapping materials. If composed of two materials, the next step decomposes the material into the two overlapping contributions. The final step is to perform recognition on the materials. To decompose overlapping materials, the authors use a method originating from Dual-Energy X-ray Absorptiometry (DEXA), which is used for measuring bone mineral density and soft-tissue composition of human bodies. The method uses quadratic approximations of the polychromatic transparencies of the high and low energy images. The authors test the algorithm on synthesised data and real data captured in a lab experiment, and achieve good qualitative results.

Other researchers have used simulations to investigate the possibility of material discrimination on systems that are not dual-energy. For example, Gil et al. [93] use Monte Carlo simulation to investigate the possibility of *single-shot material discrimination*. The single-shot method assumes that the detectors can measure the energy spectrum of the beam and can split it into a low and high energy component to determine the R-ratio. The authors simulate a Bremsstrahlung beam with 9 MeV cut-off, and a low-high division chosen at 4 MeV. They compare the one-shot R-ratio to a 4/9 MeV dual-energy simulation. Comparing the R-ratio for silver and tissue-equivalent plastic, it appears that the one-shot method has a greater discriminative effect, whilst having the potential benefits of lower X-ray dose and higher scan rate. However, it is unclear whether the system model results in a realistic level of noise when compared to a commercial system.

Fantidis et al. [103] investigate potential mixed γ- and X-ray system architectures, and their ability to discriminate materials, through Monte Carlo simulation. They simulate three γ sources (^{60}Co, ^{137}Cs, and ^{88}Y), and a

4/9 MeV dual-energy Bremsstrahlung beam. They test material discrimination performance on 165 materials and using different dual, triple and quadruple combinations of the sources. They assess the potential performance of the system using the number of R-overlaps between different materials. They claim that the optimal selection of sources are 4 MeV Bremsstrahlung and ^{137}Cs for dual, and 4/9 MeV Bremsstrahlung and ^{137}Cs for triple. The optimal quadruple source system, although not specified, offers only a slight improvement over the optimal triple source system. There is no evidence that the authors attempt to model system noise and its effects on discrimination performance. Other researchers have found that R-values alone are not a good indicator of performance due to noise in the R-estimates when interrogating materials with small or large mass thickness [5].

4.2.5 *Segmentation for CT baggage*

For CT imagery of baggage, there have been several proposals for single- and dual- energy segmentation, with some based on ML, which are reviewed here. The algorithms are designed for segmenting 3D volumes, but aspects of the approaches may be transferable to 2D cargo. In CT baggage segmentation, algorithms must cope with a variable and unknown number of baggage items, each with a wide range of possible shapes and sizes [104]. This is in contrast to the medical domain, where segmentation tasks are pre-specified, for example a segmentation of a particular organ [104]. Therefore, baggage researchers have looked to design unsupervised algorithms that make no assumptions on the number of objects or on their composition.

The approach taken by Grady et al. [104] for single-energy CT, first identifies object voxels, then identifies candidate object splits using the Isoperimetric Distance Tree (IDT) method [105], and finally evaluates good splits according to a novel Automatic QUality Assessment (AQUA) metric learnt from a large training set. The initial coarse segmentation uses a Mumford-Shah based method [106] applied to a pre-processed (denoised and artefact reduced) CT image. The AQUA method is based on a 42-dimensional descriptor from the prior literature on object segmentation, which includes features based on geometry, intensity, and gradients. To learn the AQUA model, the authors use Principal Component Analysis (PCA) to reduce dimensionality, then fit a Gaussian Mixture Model (GMM) over the PCA coefficients of all the segments in the training set using Expectation-Maximisation (EM). AQUA is

IDT is a graph-based clustering algorithm which is based on an exact clustering of a minimum spanning tree relative to a minimum isoperimetry criterion.

used both to select best candidate splits, and to select the best segmentation over three different parameter settings.

Mouton and Breckon [107] introduce a material-based segmentation for low resolution Dual-Energy Computed Tomography (DECT) images representative of the aviation security environment. After pre-processing to reduce metal artefacts, the authors first perform a coarse segmentation based on the Dual-Energy Index (DEI) and connected component analysis. The DEI combines the high and low energy linear attenuation coefficients at each voxel to give a crude estimate of the material characteristics. The authors use a Random Forest (RF) model to guide the segmentation process by assessing the quality of individual object segments and the entire segmentation. For individual object segments, the trained RF model uses the same 42-dimensional descriptor proposed by Grady et al. [104]. The authors claim that using the RF approach outperforms AQUA in their aviation setting. The quality of full segmentations is assessed using the RF score of constituent objects weighted by the error in the number of segmented objects. The authors demonstrate that their approach outperforms three state-of-the-art segmentation techniques, including: IDT [105]; Symmetric Region Growing (SymRG) [108]; and 3D Flood-Fill region growing (FloodFill) [109].

In cargo, material-based segmentation is much more challenging due to overlapping of materials and objects, and the difficulty in reconstructing linear attenuation coefficients that encode material information. However, the α-curve [97, 98], R-curve [5], and H-L curve [99] methods can provide crude (more so than DEI) material information that could potentially be used to initialise coarse segmentations. Methods similar to AQUA [104] and the RF approach of Mouton and Breckon [107] could be used to identify object splits and to assess overall segmentation quality. However, it is likely that extra metrics would be required to deal with overlapping objects without *a priori* information on the number of objects overlapping or their characteristics such as thickness and material type. Methods have been proposed in multi-view baggage for layer separation that may be applicable to multi-view cargo [110]. To best knowledge, there have been no proposals for cargo, or indeed single-view baggage, that can convincingly address these issues.

4.3 IMAGE UNDERSTANDING

Automated *image understanding* tasks in cargo can be split into the themes of ACV and ATD. An overview of the most pertinent works in the literature is given in Table 4.1.

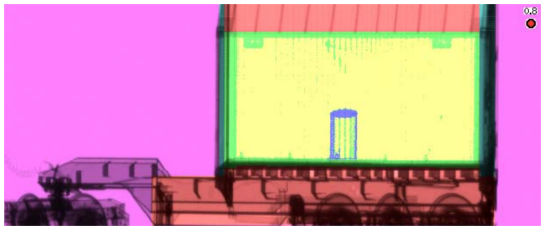
Study	Task	Methods	Notes
Chalmers [4, 70]	ECV	Intensity histogram features (min, max, and mean); compare with historical database example.	No QE given
Orphan et al. [66]	ECV	Segment floor, walls & roof; rule-based object detection	97.2% accuracy; 0.4% FPR
Andrews et al. [111]	anomaly detection & ECV	Down-sampled images; sparse auto-encoder; hidden layer features; RBF-SVM	99.2% accuracy; For features, hidden representation > normalised squared residual
Zhang et al. [6]	MV	Leung-Malik filter codebook; SIFT; dense sampling; edge sampling	visual codebook method>SIFT. Edge sampling>dense sampling.
Tuszynski et al. [112]	MV	Median intensity hist.; average absolute deviation; weighted city block distance.	48% accuracy and 5% FPR
Zheng and Elmaghraby [68]	ATD	Correlation coefficient; threshold	No QE given, detected anomalies may not correspond to presence of a threat.

Table 4.1: A summary of the literature on automated cargo *image understanding* research. Abbreviations: Quantitative Evaluation (QE); False Positive Rate (FPR); Convolutional Neural Network (CNN); Scale-Invariant Feature Transform (SIFT); Radial Basis Function Support Vector Machine (RBF-SVM). The > symbol denotes 'performs better than', and >> denotes 'performs much better than'.

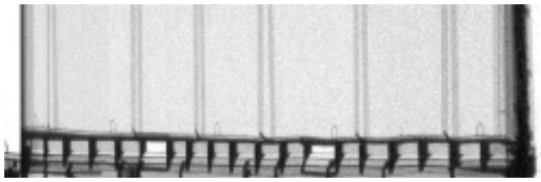
4.3.1 *Automated Contents Verification*

ACV checks whether the cargo contents match those stated on the shipment manifest. This can range from Empty Container Verification (ECV) to full Manifest Verification (MV). ECV can be useful for increasing throughput, since declared-as-empty cargoes (20% of all containers) can be sent through a separate automated inspection lane. ECV examples are given in Figure 4.4. Containers may be falsely declared as empty in shipping fraud, or may be

exploited in rip-on/rip-off smuggling operations. False declared-as-empty cargo containers can also pose safety hazards during container stacking at ports due to the unexpected additional weight. MV compares the X-ray image to the Harmonized System (HS) codes declared on the manifest. Each HS code defines a different broad category of cargo type, for example, live animals, animal products or vegetable products (Section 2.2.3).



(a) Example of ECV detection



(b) Example of ECV false positive

Figure 4.4: Example ECV results of a rule-based algorithm. The colouring in image (a) shows empty space (purple), roof and chassis (red), cargo walls (green) and an object (blue). Image (b) shows a false positive caused by the vertical spars. Sources: Orphan et al. [66].

The first work on ECV was Chalmers [4, 70], who uses 'readily available' algorithms to segment the container region and compute metrics that are then compared with empty containers of the same size. The authors give no specific details of the algorithms or their performance, but we interpret it as follows. The container is classified by generating an intensity histogram of the segmented cargo region and comparing to histograms from historical empty images. The comparison is made using histogram features such as minimum, maximum, mean, and standard deviation. Another method is briefly described by Orphan et al. [66], which segments the image (e.g. floor, walls, and roof) and then applies an unspecified rule-based object detection algorithm. The authors report 97.2% accuracy (with 0.4% false negatives) when classifying SoC images as empty or non-empty.

Andrews et al. [111] was published subsequent to the work presented in Chapter 7.

Andrews et al. [111] have recently used ECV as a test problem for anomaly detection using auto-encoders. They use cargo X-ray images of empty and non-empty containers down-sampled to 32×9 Px. In anomaly detection the algorithm is trained on the normal class only. In one test they considered empty containers as normal, and the non-empty as anomalies; in another

test, the classes were reversed. The authors derive a number of features from the hidden layer of a trained sparse auto-encoder, including: the hidden representation, the scalar residual magnitude; the signed residual (with and without normalisation by the root-mean-squared residual); the absolute residual; and the squared residual (with and without normalisation by the mean-squared residual). The features are classified using a one-class RBF-SVM. When considering non-empty containers as the normal class, they find that the one-class RBF-SVM achieves best classification accuracy (92.99%) when fed the hidden representation as a feature. When considering empty containers as the normal class the best accuracy (99.2%) is achieved when the normalised squared residual is used as the feature. No attempt is made to extend the method to detect much smaller adversarial loads, where a much higher image resolution than 32×9 Px would be required.

There have been two published attempts at MV [6, 112]. MV is a multi-class classification task, where cargo containers are classified according to HS code. Tuszynski et al. [112] used the median grey-level image histogram of each HS code in a training set. They then use a weighted city block distance to compare a given example to each HS code model. This approach yields an overall accuracy of 48% given a false positive rate of 5%. This result is improved slightly by Zhang et al. [6], who use a Leung-Malik filter bank to construct a visual codebook as a texture descriptor. They determine that this outperforms SIFT when classifying cargo images according to their HS code. Note that the authors ignore 'non-classical' examples, which they define as those containers that are less than half filled with cargo. For real-life deployable system, such examples should be included since an adversary could purposefully choose to only half fill a container when smuggling or avoiding duties.

4.3.2 *Automated Threat Detection*

Currently, much more work has been done on ATD for baggage than cargo and vehicles. Both fields are discussed here.

Zheng and Elmaghraby [68] propose a method for ATD in vehicles by detecting anomalous regions within images. They use backscatter images (one top-view and two side-views) and a transmission image (side-view) captured from an AS&E OmniView® Gantry (Figure 4.1). They perform a window-wise correlation analysis comparing a fresh image of the vehicle to

a historical image of the same vehicle stored in database. Images are split into 64 rectangular 4×16 Px windows, and the correlation between windows in similar positions in the analysed and historical images are computed, resulting in a 64×64 matrix of window correlation values. A given window is deemed anomalous if the maximum of the corresponding matrix row is below a threshold. No quantitative evaluation of the performance is given. A criticism of this proposed method is that an anomalous region will very rarely indicate an actual threat and so the false positive rate is likely to be extremely high.

More ATD research has been carried out in baggage, and detailed summaries can be found in the review by Mouton and Breckon [113]. A brief overview of the points relevant to cargo, is now given.

Several different X-ray imaging modalities are used in baggage screening. These range from single-view [114], to multi-view [92, 115–118], to full 3D CT [119–122]. Classification performance typically improves from single-view to CT as more information becomes available. The challenge is how to best use this information.

The general consensus amongst the baggage image analysis community, is that classification based on X-ray image data is more challenging than visible spectrum data, and that direct application of methods frequently used in natural images, such as SIFT, Rotation Invariant Feature Transform (RIFT), and Histogram of Oriented Gradients (HOG), do not perform well [123]. However, the performance can be improved by utilising the characteristics of X-ray baggage images. For example researchers have found that object detection can be improved by augmenting multiple views, using a 'false colour material image' [124] or using simple descriptors such as Density Histogram (DH) or Density Gradient Histogram (DGH) [119, 121] in the case of CT.

False colour, in this case, colours pixels according to the type of material i.e. material discrimination.

While it has been widely reported that texture descriptors in baggage scans perform poorly due the lack of texture in X-ray examples [123–125], the texture visible in cargo X-ray images *does* differ significantly between images. Medium to low density cargo (e.g. tyres or machinery) often contain a lot of complex articulated texture, while high density cargo (e.g. barrels of oil or bulk coal) has a more uniform appearance. This is possibly why researchers in cargo have enjoyed some success with texture descriptors such as visual codebooks based on a Leung-Malik filter bank [6].

oriented Basic Image Features (oBIFs), a good feature for texture description, also worked well for detection of cars in cargo containers in a co-authored, concurrent work [20].

Franzel et al. [118] propose a method of fusing detection results from multiple single views to exploit the extra information available from multi-view. They use a voting-based scheme where detection confidence is increased if rays from detection points from single views intersect in 3D. Their rationale is that it suppresses false alarms that do not coincide in different views, whilst reinforcing detections that do. The detection confidence on the individual single-view images are determined by sliding a window over the image, computing HOG as features and using a linear Support Vector Machine (SVM). They address in-plane rotations using a non-maximum suppression scheme, since HOG features are not rotation invariant. Moreover, the authors claim that the multi-view voting fusion scheme handles out-of-plane rotations. They achieve significantly better detection with their multi-view scheme (80%) over single-view (50%) for a 50% false alarm rate.

Multi-view fusion approaches similar to those proposed by Baştan et al. [92] and Franzel et al. [118] might be applicable to multi-view fusion in cargo, however performance is likely to be far worse due to the additional complexity. Perhaps, an approach to multi-view detection, for both baggage and cargo, would be to feed the different views into a CNN as separate channels or separate streams. The CNN can learn to jointly use information from the separate views to make better classifications. For 3D shape recognition, Su et al. [126] have found that CNNs fed with multiple 2D views as inputs performs better than using state-of-the-art 3D shape descriptors. It would be an interesting study for ATD in CT, particularly if better performance can be obtained without having to reconstruct the full 3D baggage image.

Since most approaches in baggage operate on the RGB material discrimination image, detection methods are, in some sense, utilising material information if they utilise the colour information. Baştan et al. [92] apply ATD to both the RGB image and the raw high and low energy measurements. They show that the additional information from dual-energy images significantly improves object recognition. The authors, experiment with dual-energy variants of the SIFT and domain SPIN image descriptor (SPIN) descriptors. They compute Colour SIFT (CSIFT) and Colour SPIN (CSPIN) descriptors, which operate on the individual colour channels of the RGB image. In addition, they compute Energy SPIN (ESPIN) descriptors which are computed directly on the raw high and low energy images. They found both ESPIN and CSPIN performed better than using SIFT or SPIN alone, with CSPIN achieving best performance.

Recently, Açkay et al. [127] have applied deep CNNs to ATD in dual-energy RGB baggage imagery. They recognise that there is a problem with training CNNs from scratch due to the limited availability of data. Thus they adopt a transfer-learning approach by taking a CNN, pre-trained for general image classification tasks (ImageNet [128]), and then fine-tune it for ATD in X-ray baggage. The pre-trained CNN follows the architecture introduced by Krizhevsky et al. [90], consisting of 5 convolutional layers and 3 fully-connected layers. The authors re-use the generalised feature extraction and representation in the lower layers of the CNN, whilst fine-tuning the upper layers. This achieves 99.26% detection and 4.08% false positives, which significantly outperforms prior work in the field. The authors do not comment on the possibility of training a CNN, from scratch, on data augmented with TIP imagery and realistic variation.

For baggage imagery, it appears that utilising dual-energy, either by RGB image or the raw high and low energy images, can boost ATD performance. To best knowledge, there have been no prior publications on ATD using dual-energy cargo images, or using CNNs directly on the raw high and low energy X-ray images, in any field.

4.4 DISCUSSION

Automated Analysis of cargo X-ray imagery is still a relatively young field. Over the last decade, more attention has been paid to aviation image analysis (such as baggage), since problems are generally more tractable, and because there has been more funding directed towards aviation, possibly due to the more perceivable immediate threat from terrorism. Typically, most work in cargo has been kept in-house by industry for commercial and security reasons. However, academics are beginning to form relationships with industry partners, gaining access to large image datasets with which to work. This includes the relationship with Rapiscan Systems for the work in this thesis.

In comparison to natural images, cargo X-ray images offer an interesting and difficult challenge for researchers, since objects are translucent making occlusions difficult to disentangle, are usually very cluttered and noisy, whilst appearing skewed in perspective due to the geometry of the X-ray beam. Furthermore, image contents are often more varied than images from the baggage or medical X-ray imaging domains, since a very diverse range

of objects are shipped inside containers. It is possible that more researchers would become involved in the field if data was easier to obtain, for example, through the creation of large, labelled, and open datasets.

The general development pattern of research in the field follows a similar pattern to that found in mainstream computer vision. The patten in mainstream computer vision [129] has come in three phases:

(i) algorithms completely hand-crafted by experimentation and intuition;

(ii) features hand-crafted based on intuition and experimentation, with the features classified using ML techniques;

(iii) the features and their classification learnt directly from data.

In cargo image analysis, a similar pattern is observed, however it has not yet reached stage (iii). For example, the initial works relied heavily on hand-crafting rule-based algorithms in the case of ACV, or hand-crafting algorithms based on the principles of X-ray physics in the case of material discrimination. In recent years, there are signs that a number of researchers are beginning to apply ML techniques to such problems. It is apparent that most work in baggage belongs to stage (ii), and more recently stage (iii) has been reached in the application CNNs for baggage ATD [127].

Each of the technical chapters of thesis are now discussed in the context of the existing literature.

4.4.1 *Detector wobble*

To best knowledge, no attempts have been made to address for the effects of detector wobble in X-ray cargo imagery. However, wobble does effect other imaging modalities such as C-arm CT and micro-CT, and there have been several publications on the correction of wobble in these modalities. Such methods can be characterised as being:

- *Calibration-based* – in systems where wobble is a systematic artefact that effects each images in the same way, the system or an algorithm can be calibrated to compensate for wobble;
- *Image-based* – if wobble is not a systematic artefact but a type of noise that effects each image in a different way, algorithms can be developed that operate only on the image in order to correct for the effects of wobble;

- *Online* – if an image-based approach cannot achieve then wobble is measured and corrected online.

Due to the unpredictable component of wobble (e.g. wind, uneven ground surface topology, or vibrations) in large-scale transmission radiography, it is not possible to correct wobble purely by calibration. Image-based methods (without using dedicated detectors or prior knowledge about large-scale radiography), such as TV denoising or TIWS [130], may be applicable, however they are difficult to use in practice without prior information on the severity of the wobble artefact which is measured *online* in Chapter 5. In this contribution, both a *calibration* procedure and an *online* method are employed. The *calibration* procedure is used to estimate a number of parameters that are fixed for a given system, including: misalignments of imaging sensors; the collimated width of the fan-beam; and the sensitivities of individual sensors due to housing attenuation and their intrinsic response. The *online* component deals with the estimation of wobble and estimation of the fluctuation in the photon flux, which can both vary stochastically during a traverse mode scan.

4.4.2 *Threat Image Projection*

Machine learning is a class of computer algorithms that have the ability to learn from data to tackle a task, without being explicitly programmed. In computer vision, the amount of machine learning has increased through time, both in terms of the numbers of learnt model parameters, and the fraction of the typical computer vision processing pipeline that involves learning. For example, initially, typical processing pipelines involved no learning and relied on explicit programming to tackle a task. Later, features were extracted from images using explicit programming and these features were classified using machine learning algorithms. Recently, with deep learning, state-of-the-art methods use machine learning at every step in the processing pipeline, from learning feature extraction, to learning the mapping of these features to an output.

A recent trend in mainstream computer vision, is that performance has improved as more (machine) learning is involved. Initially, performance was improved by learning the classification based on fixed image features, later performance was further improved learning the features as well as the classification. Typically, as the learnt features increase in abstraction and com-

plexity, performance too improves. However, with each improvement, larger and larger datasets are required. As such, computer vision researchers often employ data augmentation, to boost the number of training examples when data is limited. Data augmentation involves applying realistic transformations to images, which preserve the class of the image. The aim is to learn feature representations that are invariant to such transformations to improve the generalisation ability of the system. Such a method can overcome the *data problem*.

The *data problem* is particularly severe in security domains, since data on real threat items is rare, and collecting large amounts of data on staged threats is infeasible due to cost. The *data problem* has also affected human operators, for example operators not exposed frequently to threats are less likely to detect them when they appear [131]. A solution to this problem with humans was the introduction of TIP (Section 4.2.3), which boosts operator performance by exposing them to more threats on-the-job or during training, and is also used to assess operator performance. However, such methods are not known to be in widespread use in cargo.

In this thesis, a simple method for performing TIP in cargo is proposed and validated. The purpose of this method is to train and evaluate the performance of ML-based algorithms rather than humans operators. Although the methods could be equally applied to training and evaluating humans, as is conventional within current aviation security screening operations.

4.4.3 *Empty Container Verification*

Verification that declared-as-empty containers are actually empty is an important problem in terms of security and throughput. Criminals frequently attempt to smuggle inside empty containers, because it is cheap and convenient. Since, at any one time, 20% of the deployed container fleet are declared as empty, an accurate ECV algorithm operating on those 20% can potentially detect any type of crime, including shipment fraud and the smuggling of any item. A specific threat detector does not have to be developed; only an algorithm that can confirm whether the container is truly empty.

ECV is non-trivial due to the large variation in both the class of empty containers and possible loads, as discussed further in Chapter 7. Therefore simple methods such as correlation and fourier spectral analysis do not work well.

To date a few methods have been proposed for ECV. Initially these were rule-based [66] and the algorithm not fully documented or assessed (possibly due to commercial reasons). There has been one attempt [111], which applied unsupervised ML methods by treating ECV as an anomaly detection problem. However, the images used in this work were massively down-sampled from 2600×850 Px to 32×9 Px, and it is unlikely to work on small amounts of contraband smuggled inside cargo containers, since most information is lost in the down-sampling.

In this thesis, a supervised ML-based method is proposed which operates on full-sized images, and the method is tested on (i) SoC imagery for comparison with prior works, and (ii) small adversarial loads which are synthetically concealed using TIP.

4.4.4 *Dual-energy Automated Threat Detection*

Of the cargo image analysis topics reviewed, the majority have focussed on material discrimination. These methods are typically derived from physics, and no ML techniques (similar to those used in baggage [104, 107]) have been applied to the subject due to the difficulty of obtaining sufficient data with accurate labelling. Furthermore, each author tends to use a different dataset, making it difficult to compare the performance between different contributions. And most authors choose to evaluate performance qualitatively rather than quantitatively, often using only a single image.

No work in cargo has attempted to exploit material information encoded in dual-energy images to improve the performance of automated *image understanding*, including for ATD. Similar research has been performed in baggage either by operating on the material-coloured RGB image or the raw dual-energy measurements, where it was found that material information can improve ATD. In this thesis, the best way to utilise raw dual-energy measurements is investigated and a direct comparison is made between the three material discrimination methods typically used in cargo. This is the first time that the three methods have been compared, and the first time that dual-energy have been utilised to improve ATD performance. In addition, to best knowledge, it is the first time that CNNs have been applied to explicitly operate on two X-ray images measured at different energies, in any field.

4.4.5 *Concurrent Related Work*

There were several related pieces of research that were performed at the same time as this thesis, and provided important input to this work. These are summarised in this section.

4.4.5.1 *Concealed car detection*

The detection of cars in cargo containers is a good test case for the feasibility of computer vision techniques on cargo X-ray imagery. Cars typically occupy a large proportion of the container and are there easier to detect in comparison to small threats in ATD. However, concealed cars can still be difficult for operators to detect when quickly flicking through large numbers of container thumbnails. Several different Bag-of-Words (BoW) approaches [19, 20] were applied to automated car detection, including:

- Intensity histograms - a feature similar to DH used by Flitton et al. [119, 121] in baggage, histograms of both the raw and log intensity were tested;
- oBIFs - encode textural information by classifying image pixels into one of 23 categories according to local symmetry and their quantised orientation;
- Pyramid Histograms of Visual Words (PHOW) - a multi-scale extension of dense SIFT and regarded as a state-of-the-art BoW approach.

Two machine learning algorithms were trialled for classification of BoW features: RF and SVM. In addition to BoW, CNNs were also employed [19] including (i) a CNN pre-trained on the ImageNet dataset and fine-tuned for car detection, and (ii) a 19-layer very deep CNNs [132] trained-from-scratch on X-ray cargo imagery. The main conclusions of this work were that:

- RF classification based on oBIF features provides significantly better performance than using PHOW;
- Log transforming images improved the performance of the intensity histogram method and CNNs;
- The best (trained-from-scratch) CNN approach yielded a modest, but statistically significant, improvement over the oBIF approach.

4.4.5.2 *Single-energy Automated Threat Detection*

The ATD network was the same architecture as the car detection network, and based on the very deep networks of Simonyan and Zisserman [132].

Whilst, for car detection, CNNs provided only a modest performance improvement over oBIFs, for ATD the improvement was substantial. The CNN was trained-from-scratch using the TIP and data augmentation framework proposed in Chapter 6. The network provided 90% of threats with a 6% false positive rate. It was found that using a two-channel network with the raw image as one channel and the image logarithm as the second channel achieved significantly better performance than using just the raw image in a single-channel network.

4.4.5.3 *Anomaly detection*

Another way to overcome the *data problem* is to treat threat detection as a one-class anomaly detection problem. In this, no threat data is required for training, since the system is trained on the *normal class* only. The advantage is that the system can detect never-before-seen threats since the system does not need to be trained on them. However, accurately modelling the normal class is difficult and an anomaly detection system suffers from high false positive rates when used for threat detection.

Andrews et al. [21] relies on the use of (i) representation learning, and (ii) anomaly detection based on the learnt representation using a Forest of Random Split Trees (FRST). A number of different representations are tested including: (i) a baseline representation using intensity histograms; (ii) a CNN representation learnt on natural imagery from the ImageNet dataset; (iii) a CNN representation from the car detection work [19]; and (iv) a representation learnt using a Siamese CNN.

The Siamese network was trained on classifying whether two input image patches belonged to the same container or not. In this way, it was reasoned, it could build a good representation of the contents of cargo X-ray images.

The work shows that best representation for the purposes of anomaly detection was that obtained by the Siamese network, which yielded an Area Under the Curve (AUC) of 63.7% when the system was tested on an ATD task. Moreover, separate testing of the FRST algorithm showed that it was suitable for anomaly detection given a suitable representation with which to work. For a deployable anomaly detection system, however, the representation learning needs to be improved to obtain results competitive with supervised ATD approaches. Anomaly detection is not pursued as a topic in this thesis.

4.5 SUMMARY

In this chapter the state-of-the-art for automated cargo image analysis, prior to this thesis, was explored and discussed. During this analysis, it was found that research in cargo has not kept up to speed with other application areas of image analysis, such as aviation baggage image analysis or medical image analysis. The majority of algorithms for cargo are still based on intuiton- or physics-derived methods, whereas in other fields machine learning based approaches have become widely adopted and shown to offer much greater performance. This phenomenon is probably due to the difficulty to obtain sufficiently large datasets for training machine learning based algorithms.

In the rest of this thesis, we address four problems identified in this chapter:

- In Chapter 5, the problem of detector wobble artefacts is addressed using a machine learning based wobble estimation algorithm, and an image correction algorithm based on these wobble estimates, which is derived from the principles of X-ray physics introduced in Chapter 3. There have been no prior publications on wobble correction in cargo images, however there has been analogous work performed for other imaging systems such as C-arm CT.

- In Chapter 6 we introduce a TIP framework for training and testing machine learning based image understanding algorithms. The aim of this is to overcome the problem of limited datasets for developing such algorithms. The TIP framework is used in the final two technical chapters to train machine learning based algorithms. A potential limitation of training on TIP imagery is that the algorithm will overfit to TIP imagery if there are cues present in TIP imagery that are not present in real imagery. To this end, a number of methods (including Empty Image Projection (EIP)) are proposed to verify that this does not occur.

- In Chapter 7, the problem of ECV is addressed using a machine learning approach. From the prior art, there have been two research groups attempting to tackle ECV. Both of these have not used machine learning, can only detect large cargo container loads, and provide very little evaluation of the methods. In Chapter 7, it is shown that a machine learning based approach is able to outperform these previous methods for large loads, and is also extendible to the detection of very

small loads, such as a few kilogrammes of cocaine smuggled in an otherwise empty container (rip-on-/rip-off). This is an important step since empty containers constitute 20% of the container fleet, and are often left unlocked, giving criminals easy access.

- In Chapter 8, a deep learning based ATD algorithm is developed which operates on dual-energy cargo images. This is the first use of dual-energy images in cargo ATD algorithms in the literature, with the motivation being that it can help to reduce false alarms by implicitely learning to exploit material properties as encoded by the raw dual-energy measurments. The algorithm significantly outperforms prior work of Jaccard et al. [10].

5

MEASUREMENT AND CORRECTION OF DETECTOR WOBBLE

Here, we address the affect of detector wobble on large-scale transmission X-ray imaging. A series of corrections for wobble artefacts and noise are derived by forming a model of image formation in the presence of wobble. A method of measuring wobble is proposed which can be used in image correction. The measurement relies on the use of Beam Position Detectors (BPDs), which are imaging detectors placed perpendicular to the imaging array. The methods are tested both qualitatively and quantitatively on real images captured by a system modified by rotating four imaging detectors to create BPDs.

5.1 MOTIVATION

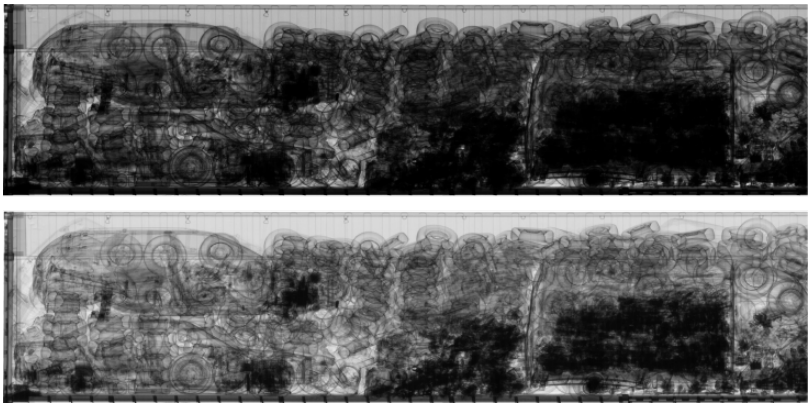


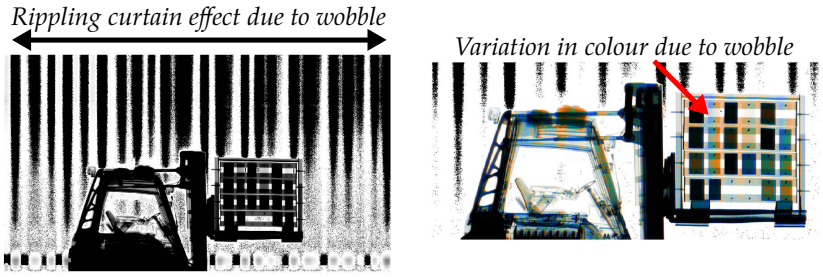
Figure 5.1: A raw transmission X-ray image of a cargo container containing vehicles and vehicle parts (top) and an intensity manipulated version (bottom). Intensity manipulation is often used to reveal details in the image when searching for threats.

For operators to quickly and accurately detect threats, high spatial resolution and accurate image values are required [133]. The former, because threats may be small, and the latter because threats may be shielded by other cargo or only revealed by subtle differential absorption. State-of-the-art transmission systems offer imaging of vehicle contents at resolutions of

a few mm/pixel [29] and precisions of 16 bits. In some systems, mechanical instability (wobble) leads to effective loss of precision.

Whilst large-scale X-ray Computed Tomography (CT) could alleviate the issues of wobble and shielding, such systems are not widely deployed because they are too expensive and inefficient to be competitive [51, 134].

In Chapter 3, the concept of detector wobble was introduced and how it can lead to artefacts, noise or geometric distortions depending on the type of system used and the severity of the wobble. This chapter focuses on induced noise and artefacts due to wobble; this presents in the image as a variation in pixel intensity (Figure 5.2). In systems where the effect is systematic between scans, wobble is an artefact. In systems where the effect varies between scans due to stochastic factors, wobble is noise. Only traverse mode systems are effected by detector wobble.



(a) Raw image (b) Material discrimination

Figure 5.2: An X-ray image of a fork-lift truck from a mobile scanner with mechanical instability (left) and the same image with material discrimination applied (right). The image grey-levels have been clipped to make visible the small changes in image value due to wobble, which leads to a rippling curtain effect across the image. Each rectangular test piece corresponds to a single material of uniform thickness. The wobble artefact affects the classification of material type; the classification of a single test piece can change from plastic through to steel due to wobble. This results in a colour change across the test piece (indicated by red arrow) in the material discrimination image, where there should be no change.

In the traverse mode, the imaging array may wobble as it moves across the scene due to uneven ground or vibrations from the engine (truck systems), oscillations in the boom (truck and rail systems), or due to wind or vibrations from traffic (truck and rail systems). This has a particular impact when operators search for threats placed in dense scenes, since under intensity clipping [135] the wobble artefact becomes apparent (Figure 5.2). Furthermore, wobble reduces the quality of material discrimination images [5] since their computation is dependent on precise values. Discrimination of high atomic numbers is particularly important as it can reveal smuggled nuclear materials, or their shielding [66, 136]. Wobble occurs in both truck-mounted and gantry systems. In truck-mounted systems wobble is variable from scan to scan, but in gantry systems it is systematic. In this work a gantry system is employed, since it allows determination of the wobble ground truth, but

methods developed herein can equally be applied to truck-mounted systems.

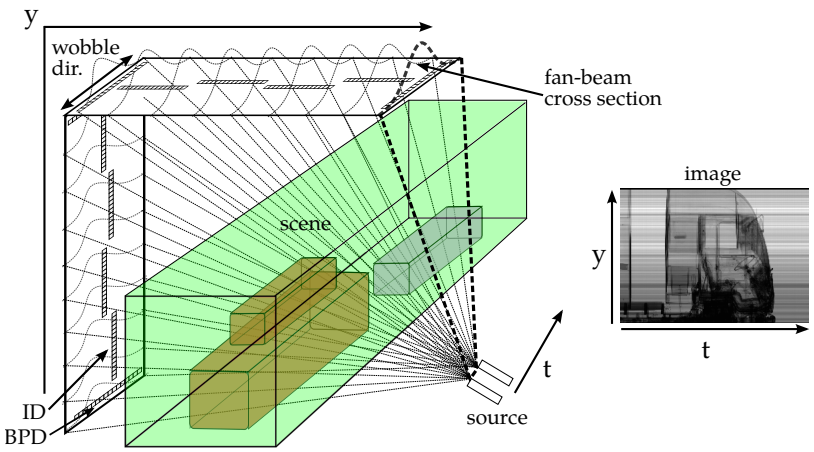


Figure 5.3: An illustration of the experimental set-up employed in this work, including the placement of four Beam Position Detectors (BPDs), created by rotating imaging detectors by 90° relative to the imaging detectors. BPDs allow the intensity profile across the beam width to be measured. Translation of the scene relative to the source and detector produces image columns, whilst each image row corresponds to a single sensor position in the imaging array.

In this work it is proposed that wobble can be measured using BPDs which are placed perpendicular to the imaging array as in Figure 5.3. The BPD can measure the profile of the X-ray fan-beam across its width, thus allowing for determination of the beam centre. By tracking the movement of the beam centre at different points along the imaging array, and relative to it, it is possible to track detector wobble during a scan. These wobble estimates can be applied to image correction. But to do so, a model of image formation in the presence of detector wobble should be deduced.

5.2 A MODEL OF IMAGE FORMATION WITH WOBBLE

To describe image formation with a wobbling detector, three coordinate systems are defined (Figure 5.4):

- $y \in \mathbb{Y}$ – the coordinates of imaging sensor pixels along the Γ-shaped imaging array (image vertical);
- $t \in \mathbb{T}$ – the time coordinates indexing each scanning moment during image acquisition (image horizontal);

- $x \in \mathbb{X}$ – the coordinates along the orientation of the BPDs (perpendicular to the beam and imaging array). The origin x=0 is taken as the vertical mid-line (dashed in Figure 5.4) of the imaging array.

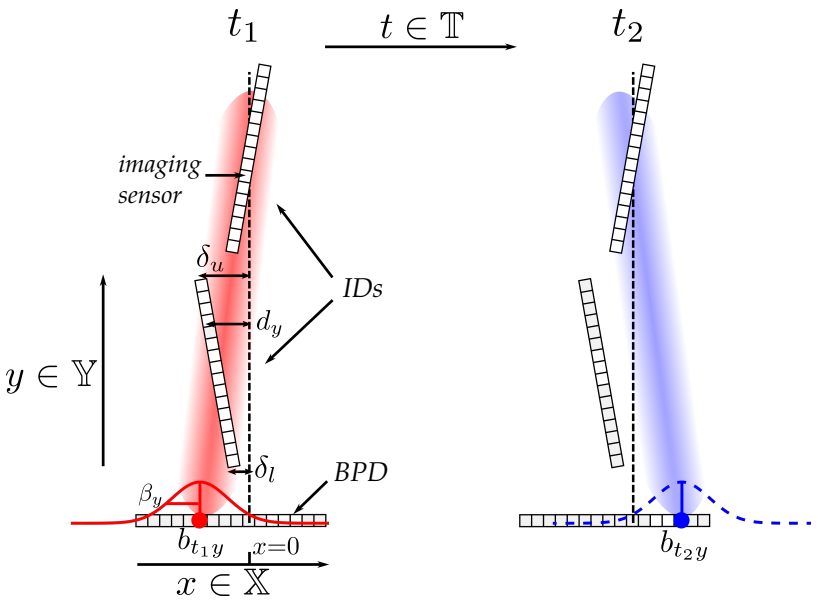


Figure 5.4: *Left*: Part of the imaging array showing two misaligned ID, a BPD, and a wobbling fan-beam. The magnitude of the wobble and the sensor misalignments have been exaggerated in this figure. The offsets d_y for individual imaging sensors are confined to a linear function determined by the offsets $\{\delta_l, \delta_u\}$ of the ID endpoints. The fuzzy bars illustrate the fan-beam incident on the imaging array. The Gaussian (width β_y and position b_{ty}) shows the profile of the fan-beam on the BPD. *Right*: A later time point $t_2 > t_1$. Due to wobble, the fan-beam has moved relative to the imaging array so that the intensity recorded by the IDs has changed. This leads to an effective loss of image precision. Correction requires estimation of the beam displacements b_{ty} and the offsets d_y to be estimated. The b_{ty}, dense in t and y, can be interpolated from estimates dense in t but computed at the sparse y values where BPDs are located.

The formation process of an image $I_{ty} \in \mathbb{R}^+$ is described as follows. The X-ray source emits a photon flux $A_t \in \mathbb{N}$ at scanning moment t. This flux is collimated into a fan-beam of width β_y, which has a spatial distribution on the imaging plane according to

$$\exp\left(-(b_{ty} - d_y)^2/(2\beta_y^2)\right). \tag{5.1}$$

The parameters $b_{ty} \in \mathbb{X}$ define the displacements of the beam cross section maximum from the vertical mid-line: when wobble occurs this varies with t and y, without wobble only with y. The parameters $d_y \in \mathbb{X}$ are the horizontal offsets of the imaging sensors from the vertical mid-line.

For a given linear ID with endpoint offsets $\{\delta_l, \delta_u\}$, d_y is constrained to a linear function

$$d_y := (y_u - y_l)^{-1}((y_u - y)\delta_l + (y - y_l)\delta_u), \text{ where } y_l < y < y_u. \quad (5.2)$$

It is assumed that the individual imaging sensors are perfectly aligned within the ID, but that each ID is misaligned relative to other IDs.

X-ray photons pass through the scene and interact via absorption and scattering. Let the scene transmission, the fraction of unattenuated photons, be denoted by $S_{ty} \in [0, 1]$. This is dependent on the thickness and type of material composing the scene. The final measured image is determined according to a sensitivity factor $R_y \in [0, 1]$, which incorporates (i) the fraction of photons that are transmitted through the sensor housing and not absorbed or scattered, and (ii) the fraction of photons impinging on the detector that are counted (the intrinsic response of a sensor).

Therefore, the final image, assuming no cross-pixel effects such as photon scatter or detector cross-talk, is approximated by

$$I_{ty} = A_t \cdot \exp\left(-(b_{ty} - d_y)^2/(2\beta_y^2)\right) \cdot S_{ty} \cdot R_y. \quad (5.3)$$

This is similar to the equation of image formation presented in Section 3.3, but energy dependence has been dropped and R is similar to $D(E)$ but also includes effects from the detector housing.

The scene transmission S_{ty} is the physical quantity that one is attempting to measure, therefore the ideal image is

$$\underbrace{S_{ty}}_{\text{ideal}} = \underbrace{I_{ty}}_{\text{raw}} \cdot \underbrace{(A_t \cdot \exp\left(-(b_{ty} - d_y)^2/(2\beta_y^2)\right) \cdot R_y)^{-1}}_{\text{correction factor}}. \quad (5.4)$$

To obtain the ideal image, one must estimate the different components of the correction factor. In the *portal* scanning mode, correction is straightforward. Absence of wobble means that $b_{ty} = b_y$, so that all that needs to be dealt with is:

1. image column variations due to fluctuations of the photon source A_t;
2. image row variations due to sensitivity R_y, and the fixed position and geometry of the beam $\exp\left(-(b_y - d_y)^2/(2\beta_y^2)\right)$;
3. image pixel variations due to Poisson noise in the number of photons that reach an imaging sensor.

Note that Poisson noise can also be ameliorated by increasing the beam intensity or exposure time, but this has implications on safety and cost.

The image column and row variations (1 and 2) can be corrected by normalising the columns and rows in the image respectively. In this work, no attempt to correct Poisson noise (3) is made, however there are several denoising algorithms for Poisson-distributed noise [137, 138] in the literature.

In the *traverse* scanning mode, where wobble *does* occur, the correction is complicated. The beam position b_{ty} now varies with t as well as y, and the imaging sensor offsets d_y must now also be estimated. These and the other parameters in the correction factor, in Equation (5.4), can be separated into two classes; (i) *system parameters* (β_y, $\{\delta_l, \delta_u\}$ and R_y) that are estimated in a one-off calibration described below, and (ii) *dynamic parameters* (b_{ty}, A_t) that are estimated per time point (online). The source variation A_t is straightforward to address by taking an image patch from a single ID close to the source and averaging it over rows. In the remainder of this chapter, A_t-corrected images are assumed. In Section 5.3, a method for estimating b_{ty} is described.

In the one-off calibration, for each BPD β_y (beam width at the BPD location) and R_x (the sensitivity of the sensors along the BPD) are estimated. For each ID, $\{\delta_l, \delta_u\}$ (the misalignments of the ID at its endpoints) and R_y (the sensitivity of the sensors along the ID) are estimated. The estimates are determined by model fitting to data collected during a traverse (wobbling) scan of an air-only scene. Although wobble has a detrimental effect on image precision, one benefits from wobble in these estimations since it allows one to disentangle (i) β_y and R_x, and (ii) b_{ty} and $\{\delta_l, \delta_u\}$.

The calibration is two-step and summarised as follows. First, a Sum of Squared Errors (SSE) minimisation model fit is performed using a Gaussian model of the fan-beam incident on the BPD, masked by the sensor sensitivities. In the fitted model, the Gaussian centre is allowed to vary freely with time but the beam width and sensitivities are unvarying. Having estimated the unvarying beam widths and the time-varying beam positions at each BPD, these are linearly interpolated to the positions of the sensors of the IDs. With these estimated, next a model fit is performed to determine the ID parameters. A SSE fit is performed on the data from each ID to jointly estimate $\{\delta_l, \delta_u\}$ and R_y. The SSE is taken between the ideal image (raw image multiplied by correction factor, as in Equation (5.4)) and a uniform unit-valued image. The correction factor is composed using the interpolated β_y and b_{ty} estimates (from step 1), and the estimated parameters $\{\delta_l, \delta_u\}$ and R_y.

The ideal air image should be constant modulo Poisson noise and with pixel values close to unity.

5.3 WOBBLE ESTIMATION ALGORITHM

To estimate wobble for inhomogeneous scenes, it is required to estimate b_{ty} at the BPDs, and then interpolate between these values to obtain estimates

at all ID locations. However, the simple model fitting of the previous section is not applicable for inhomogeneous scenes. At some scanning moments the beam will be distorted from a Gaussian shape, and at other moments it will be undetectable due to dense objects in the scene. To cope with this, the beam position at time t is estimated by fusing an instantaneous estimate $\hat{b}_{inst}$ (with uncertainty $\hat{\sigma}_{inst}$), with an estimate $\hat{b}_{prior}$ (with uncertainty $\hat{\sigma}_{prior}$), based on the previous n beam position estimates.

5.3.1 *Instantaneous estimation*

The profile (D_{tx}) measured at each instant by a BPD, is a multiplicative combination of (i) the beam profile (P_{tx}), (ii) the scene transmission (S_{tx}), and (iii) the sensitivity (R_x). The beam profile is estimated from the measured profile, fixed estimates of the sensitivity (Section 5.2), and dynamic estimates of the scene transmission estimated from previous time-points of the BPD signal, according to

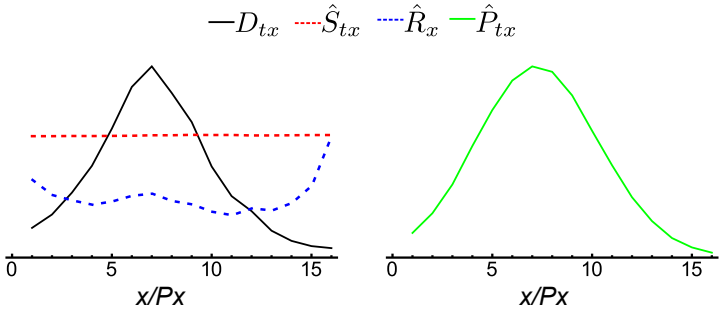
$$\hat{P}_{tx} = D_{tx}/(\hat{R}_x \hat{S}_{tx}). \tag{5.5}$$

Gaussian (and piecewise Gaussian) model fitting have also been tested for instantaneous estimation, as described more fully in [11], but it was found that the non-normal distribution of the errors makes estimation of the uncertainty unreliable. The method is not presented in this thesis.

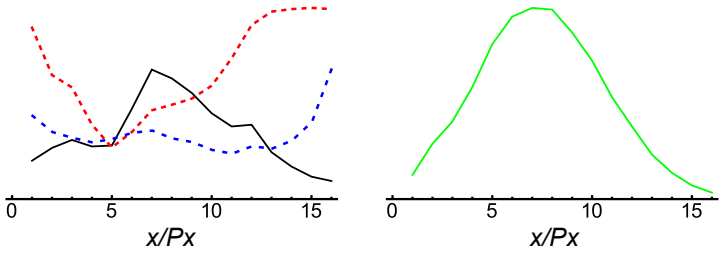
This estimation works well in cases where the scene is not too dense (Figures 5.5a and 5.5b); but when it is the estimated beam profile can be inaccurate due to (i) the low (noise-dominated) sensor signal, or (ii) deviation of photon trajectories due to scatter (Figure 5.5c).

The scene transmission function $\hat{S}_{tx}$ was estimated using measurements of the BPD as it slides across the scene. A given pixel on the BPD samples each point, at its y-value, in the scene (Figure 5.6). Plotting the response of this pixel as a function of time gives an estimate of the scene transmission function. Since each of the BPD sensors also sample each point in the scene, one can construct a similar estimate for each sensor. The final estimate of $\hat{S}_{tx}$ is obtained by taking a weighted average of the estimates from each of the sensors. The weighted average is taken to reduce noise in the estimate from sensors that are aligned with the low signal tails of the Gaussian cross section.

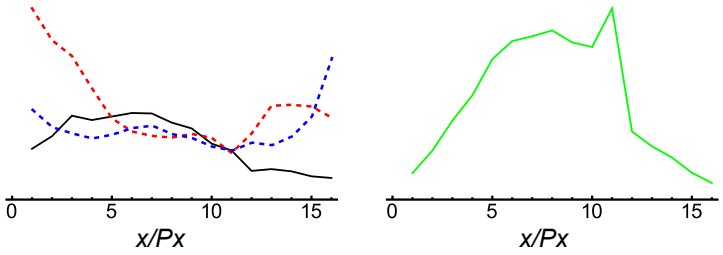
With the estimate of the beam profile ($\hat{P}_{tx}$), one can estimate the instantaneous beam position b_{inst} and its uncertainty σ_{inst}. The estimator should be able to deal with non-linear relationships in the data and be able to produce data dependent uncertainty estimates.



(a) Homogeneous scene



(b) Inhomogeneous scene



(c) Dense inhomogeneous scene

Figure 5.5: Examples of the estimated beam profile $\hat{P}_{tx}$ (green) computed by dividing the measured BPD profile D_{tx} (black) by estimates of the scene transmission $\hat{S}_{tx}$ (red) and the sensor sensitivity $\hat{R}_x$ (blue). *Top:* Example of a homogeneous scene, and thus the estimate $\hat{S}_{tx}$ is flat, resulting in a Gaussian $\hat{P}_{tx}$. *Middle:* Example of an inhomogeneous scene, the resulting $\hat{P}_{tx}$ is approximately Gaussian. *Bottom:* Example of a dense inhomogeneous scene, where the resulting $\hat{P}_{tx}$ is non-Gaussian which and is probably attribute to photon scatter.

In this work, a Random Regression Forest (RRF) [139] is used to construct a robust estimator of the beam position from the beam profile estimates. A RRF model is based on an ensemble of decision trees and is capable of modelling non-linear relationships as required. Each tree in the RRF produces an estimate of the beam position. Estimates of the instantaneous beam position $\hat{b}_{inst}$ and its uncertainty $\hat{\sigma}_{inst}$ are obtained by taking the mean and standard deviation of the tree responses, respectively. It is observed, in this study, that the standard deviation of the tree responses has a strong correlation with the actual error in the beam position estimate. Other advantages of RRF is that it is fast to train and deploy, and resistant to overfitting.

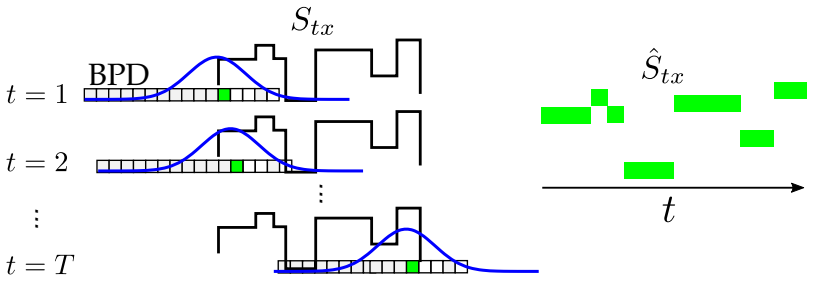


Figure 5.6: *Left:* Illustration of a BPD being translated across a scene during a scan. At consecutive time-points $t = \{1, 2, \ldots, T\}$ a given sensor (green) samples consecutive points in the scene. *Right:* plotting these samples as a function of t yields an estimate of the scene transmission. Each BPD sensor gives a similar estimate, and weighted average of is taken to reduce the noise in the final estimate $\hat{S}_{tx}$. Sensors towards the ends of the BPD, which receive low signal, are given a lower weighting in the average than those near the Gaussian centre which receive a higher signal.

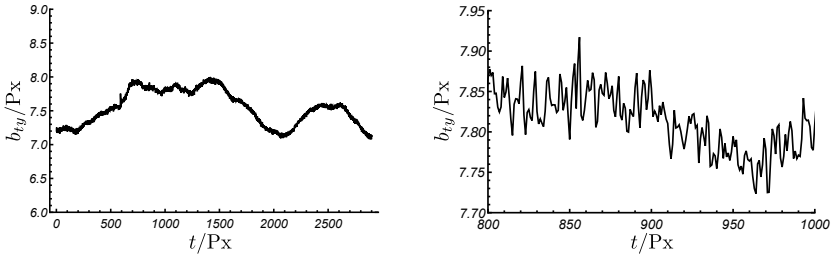
In the RRF, N_t trees are constructed top-down with bagging and random subspace sampling. Internal nodes are split using standard thresholding, and optimised according to the Residual Sum of Squares (RSS). At each split m features (i.e. scene and sensitivity normalised BPD pixels; elements of $\hat{P}_{tx}$ at fixed t) are randomly sampled. For stopping criteria, a minimum of two samples per split is enforced with no limit on tree depth. To tune N_t and m, first m is set to the recommended default (m=1/3×# features=5) for regression. N_t is then varied and the Root-Mean-Square Error (RMSE) is computed to choose a sufficient number of trees so that the RMSE is stabilised but not too many that computation time is slow. With the N_t fixed, m is then varied from 3 to 12 to find the optimal RMSE, before verifying N_t again as before. By this method, it was determined that N_t=500 was adequate and the default m=5 was optimal.

The `randomforest-matlab` implementation of RRFs [140] is employed. For training, the ground truth values of the beam displacement were obtained by use of a gantry system in traverse mode, described later in Section 5.4. Separate RRFs are trained for each BPD, using 1.4×10^5 measurements from five independent scans so that there is no overlap with the test images used in Section 5.4.

5.3.2 *Estimation based on previous estimates*

In cases where the BPD is heavily obscured, giving low Signal-to-Noise Ratio (SNR), the RRF-based instantaneous estimate will give a poor estimate of the beam position and a high uncertainty. In these cases, one wants the

beam position estimate to be sensible, and to achieve this information about prior beam positions was incorporated using an Auto-Regression (AR). The wobble of the detector array is partly deterministic (consider a swinging pendulum), but also stochastic due to the variable scanning surface, wind and vibrations. An AR is capable of learning some of the deterministic wobble, whilst allowing for stochastic variation. It is also simple to implement and fast to compute. Moreover, it is observed that the beam position trace has a low frequency component due to the wobble of the imaging array (Figure 5.7a); and a high frequency component due to either fluctuations in detector response or of the photon source (Figure 5.7b). The high frequency element makes simple estimation, based on the previous time point, unreliable. An AR, however, allows incorporation of n previous time-points, where n can be tuned on data to achieve best performance. Moreover, the AR approach effectively smooths out erroneous estimates from previous time-points, but is beneficial over other smoothing filters (e.g. median filter) since it is possible to propagate errors from previous time-points to be used in fusion (Section 5.3.3) with the instantaneous estimate.



(a) Low frequency (b) High frequency

Figure 5.7: The beam position b_{ty} as function of time t during a wobbling air scan. The low frequency component (a) is due to wobble. On zooming, a high frequency component is also visible (b).

The AR model predicts the current beam position based on a linear combination of the previous n beam positions with an added, normally distributed, perturbation

$$b_t = \sum_{t'} w_{t'} b_{t-t'} + N(0, \epsilon^2) \text{ s.t. } \sum_{t'} w_{t'} = 1, \tag{5.6}$$

where $1 \leqslant t' \leqslant n$.

The AR weights $w_{t'}$ are determined by model fitting Equation (5.6) to an independent air scan. The constraint $\sum w_{t'}=1$ ensures that the model does not have an unrealistic systematic drift. The uncertainty ϵ is determined by

applying the model to a second air-only scan and computing the RMSE. The fitted model is used to generate the prior beam position estimate and its uncertainty according to:

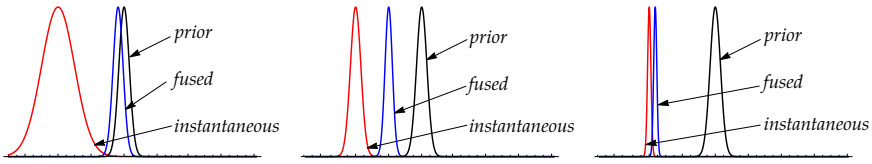
$$\hat{b}_{\text{prior}} = \sum_{t'} w_{t'} \hat{b}_{t-t'}, \;\; \hat{\sigma}^2_{\text{prior}} = \sum_{t'} w_{t'} \hat{\sigma}^2_{t+t'} + \epsilon^2. \tag{5.7}$$

Note that the uncertainties from previous time-points are propagated when forming this estimate, so that if the AR operates on previous estimates that are highly uncertain they are incorporated into the AR uncertainty, which is useful in the fusion step.

5.3.3 *Fusion of estimates*

To incorporate the information from the previous time-points, the estimates from the AR and RRF models are first fused, according to their uncertainties. The fusion should weight the final estimate more towards the AR if the RRF-based estimate is more uncertain (e.g. due to low SNR). Equally, if the AR uncertainty is high, because many of the previous n RRF-based estimates were also uncertain, but the next instantaneous estimate is very certain, then the fusion should weight more towards the RRF-based instantaneous estimate. To achieve this, a Bayesian fusion was performed. This approach is illustrated in Figure 5.8.

The Bayesian fusion is equivalent to a Kalman Filter [141].



(a) Certain prior (b) Equal certainty (c) Certain instantaneous

Figure 5.8: Demonstration of Bayesian fusion of a prior estimate (black) and an instantaneous estimate (red) to obtain a fused estimate (blue). The width of the Gaussians correspond to the uncertainty on the estimate, and their centroid to the estimate value. The x-axis can be imagined as the BPD. In (a) the prior has a higher certainty than the instantaneous estimate and so the fused estimate is weighted towards the prior, in (c) the opposite is true. In (b) both estimates have equal certainty and so the fused estimate compromises between the two.

To estimate the beam position $\hat{b}_t$ and its uncertainty $\hat{\sigma}_t$, the instantaneous estimate $\hat{b}_{\text{inst}}$ and its uncertainty $\hat{\sigma}_{\text{inst}}$ (Section 5.3.1) were Bayesian

fused with a prior estimate $\hat{b}_{prior}$ and its uncertainty (Section 5.3.2). This is expressed as:

$$\begin{aligned} \hat{b}_t &= (\hat{b}_{inst}\hat{\sigma}^2_{prior} + \hat{b}_{prior}\hat{\sigma}^2_{inst})/(\hat{\sigma}^2_{prior} + \hat{\sigma}^2_{inst}), \\ &\text{with } \hat{\sigma}^2_t = (\hat{\sigma}^2_{prior}\hat{\sigma}^2_{inst})/(\hat{\sigma}^2_{prior} + \hat{\sigma}^2_{inst}). \end{aligned} \tag{5.8}$$

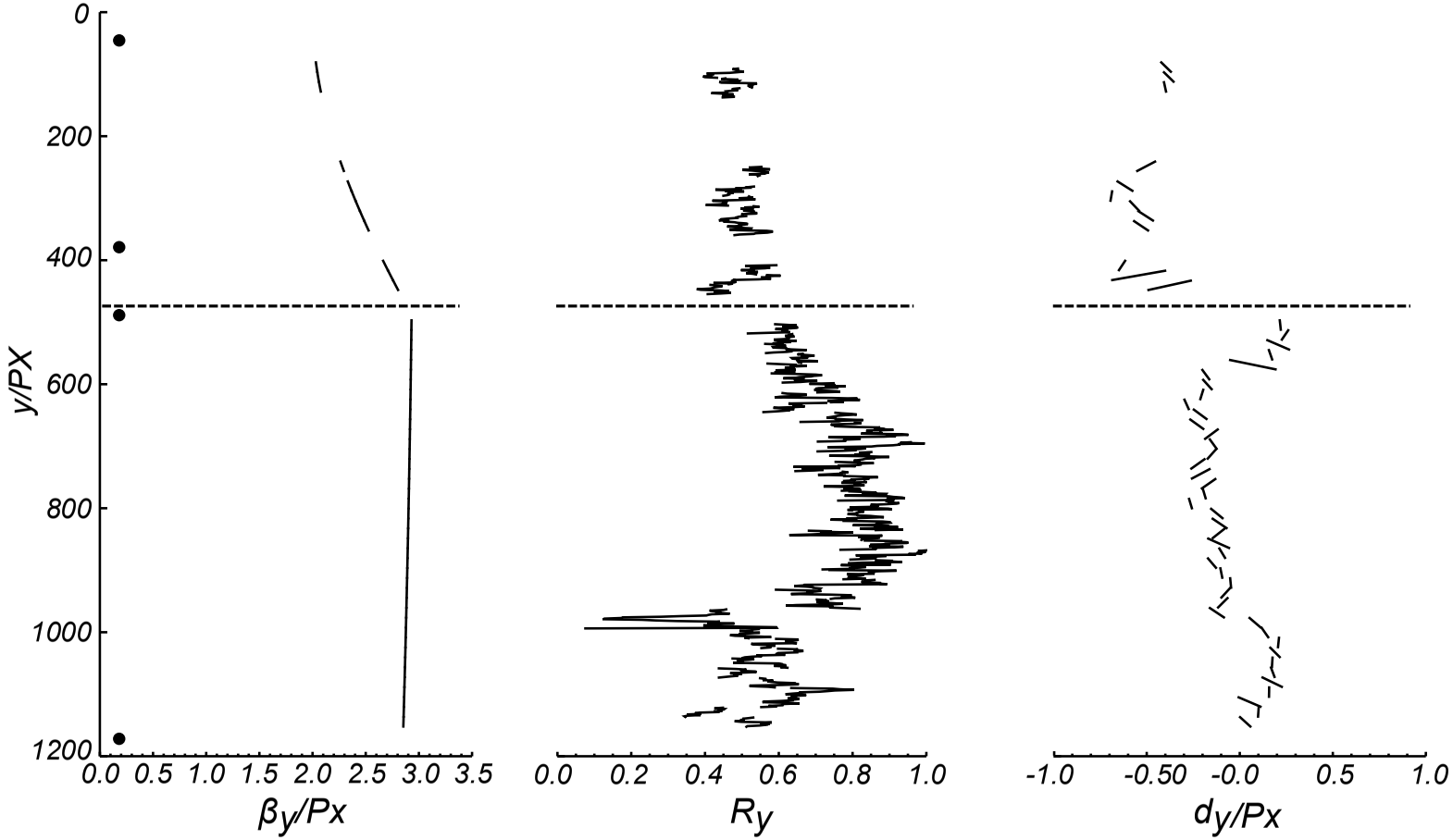
This weights the two beam position estimates by their uncertainty, as required. If the uncertainty of an estimate is low then that estimate contributes more to the fused estimate. In particular, if the instantaneous estimate is uncertain because of dense shielding, the prior estimate will be relied on; but when it is certain it will dominate the overall estimate.

5.4 RESULTS

The data was collected by James Ollier, Rapiscan Systems.

For the purposes of this study, and to test out the methods described above, data was collected using a modified Rapiscan Eagle® G60 transmission X-ray scanner. Four of the IDs were rotated by 90° to become BPDs. The BPDs were placed at the extremes of the vertical boom and the horizontal boom, so that there were two BPDs per boom. The wobble characteristics are different at each location, for example wobble is most severe at the bottom of the vertical boom. Note that in a commercial implementation of BPDs, the system would have a full set of IDs with additional detectors for BPDs, but for the purposes of these experiments this modification was adopted to reduce cost. Air-only images in portal and traverse modes were collected, and several traverse mode scans of objects (i.e. trucks, fork-lifts, scissor lifts) were performed. The scanner operates at 90 Hz and has a pixel size of 5.6 mm, giving an effective spatial resolution of approximately 3 mm. The scanner uses a Bremsstrahlung beam with a cut-off energy of 6 MeV. This is the same energy used in commercial systems, and gives enough penetration to achieve reasonable SNR on the BPD for most objects.

A gantry set-up was adopted, since it provides a ground truth for wobble. Wobble is observed in both gantry and truck-mounted systems, with a similar amplitude and frequency composition. However, for a gantry system, wobble is the same (modulo alignment) for each scan, but variable for truck-mounted systems. The gantry system allows one to obtain an accurate ground truth by aligning wobble estimates from an air-scan with the air parts of an object scan.



(a) Beam width (b) Sensor sensitivities (c) Imaging sensor offsets

Figure 5.9: Estimated system parameters: (a) beam width, β_y; (b) sensor sensitivities, R_y; and (c) horizontal imaging sensor offsets from the vertical, d_y. The dashed horizontal line marks the transition from the vertical (below) part of the Γ-shaped imaging array, to the horizontal (above). The black dots indicate the y-positions of the BPDs. Gaps in y-values are where an ID has been removed or rotated to form a BPD in the experimental set-up. The beam width increases (decreases) as the distance from source to the array increases (decreases), due to dispersion. The sensitivities fluctuate between adjacent sensors due to their different intrinsic responses. The estimated sensor offsets are piece-wise linear because they are grouped by ID, and are of the order of a few millimetres which is within the manufacturing tolerance of a system of this scale.

5.4.1 *System parameter estimation*

The system parameters β_y (beam width), R_y (sensitivities) and d_y (imaging sensor offsets) were estimated according to Section 5.2, and are shown in Figure 5.9. Small and large y-values correspond to the bottom and top of the image, or the vertical and horizontal parts of the Γ-shaped imaging array, respectively. The gaps in y-values are where an ID has been removed or rotated to form a BPD.

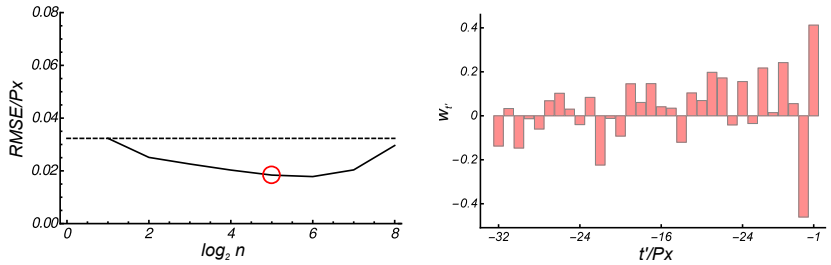
The estimate of β_y (Figure 5.9a) increases as you go along the horizontal of the Γ-shaped imaging array and away from the source due to beam dispersion; it then decreases as you go along the vertical of the Γ-shaped imaging array and slightly closer to the source. The sensitivities R_y (Figure 5.9b) have a lot of variation between adjacent imaging sensors due to their intrinsic response and due to variations in the housing of the Γ-shaped array. The estimated offsets of the IDs (Figure 5.9c) are of the order of a few mm, which when compared to their 10 cm length is plausible for a human engineer placing them during the construction of the scanner, and is indeed within the manufacturing tolerance of a scanning device of this scale (6 m tall). Note that the piecewise-linear nature of d_y is due to the linear constraint placed on each ID, as in Equation (5.2).

5.4.2 *Wobble estimation*

The AR was trained on an air-only traverse mode image. Figure 5.10a shows the RMSE performance of the trained AR on an independent air-only test image as a function of the number (n) of previous time-points considered. As n is increased the RMSE decreases, reaching a minimum at n=64, before the RMSE begins to grow. When n gets too large the model overfits and performance deteriorates. The point n=32 was chosen since the RMSE is near optimal but requires half the number of parameters. The AR weights for n=32 are shown in Figure 5.10b. It shows that more importance is placed on the most recent b estimates as expected. The oscillating structure is the AR system's way of coping with the high frequency component of the beam movement.

The Gaussian method was a direct model fit on the BPD data as in [11].

To assess the performance of the proposed beam position estimates, the performance was tested on *easy* (scissor lift, which occluded a single BPD, *intermediate* (fork-lift truck that occluded two BPDs) and *difficult* (a truck which



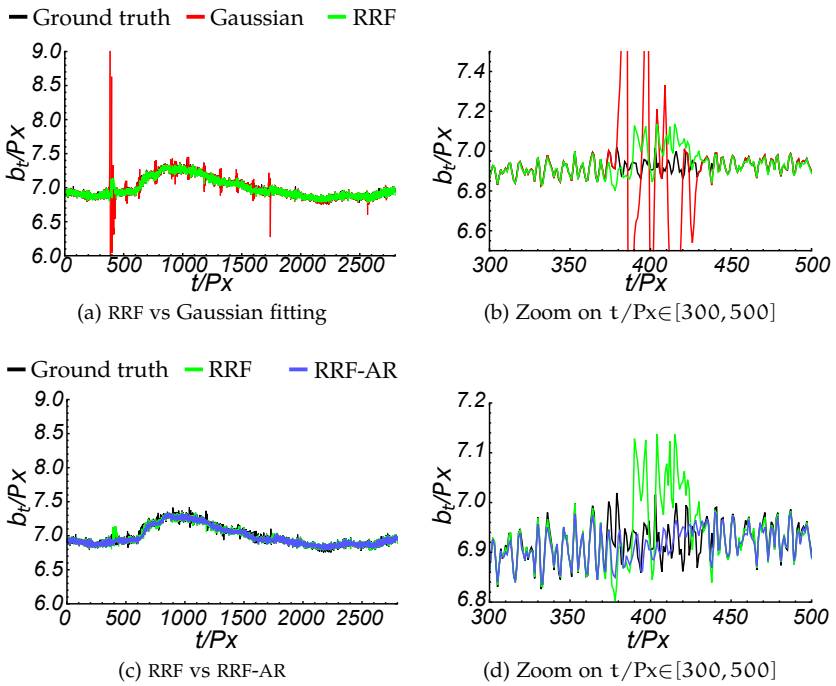
(a) RMSE as a function of n (b) Learnt weights for $n=32$

Figure 5.10: AR model fit: (a) The RMSE performance of the AR for different numbers of previous time-points n included in the model (dashed line indicates the standard deviation of the beam position, red circle indicates the near-optimal $n=32$); and (b) the learnt AR weights $w_{t'}$ when $n=32$. The RMSE decreases as the number of time points n included in the model increases, it reaches an optimum at around $n=64$, before rising again due to overfitting. The AR weights have larger magnitude for the most recent time-points ($t'=-1,-2$) as expected, because these are most informative for predicting the next beam position. The oscillating structure in the weights is the AR's way of coping with the high frequency wobble component.

occluded two BPDs throught entire scan, and contains a very dense engine region which severely limits the SNR on the BPDs) scenarios from the collected data. For each, the RRF-based method for instantaneous estimation is compared to a Gaussian-based method. Also compared are the RRF-based method instantaneous method, with the fused estimate which is referred to as RRF-AR.

For the *easy* scenario (Figure 5.11), the RRF instantaneous estimate (green) is mostly accurate, with most estimates close to the ground truth (black). The Gaussian-based method (red) gives wildly inaccurate estimates when the BPD is occluded by an object thus resulting in a non-Gaussian BPD profile. However, the RRF yields estimates much closer to the ground truth, in these cases. These estimates are made very accurate when fused with the AR (blue), since the RRF trees give variable responses which results in a larger uncertainty, so the fusion gives more weight to the AR. In particular, in Figure 5.11d, the fused estimate is much closer to the ground truth than the RRF on its own.

In the *intermediate* scenario (Figure 5.12). The Gaussian-based method does even worse, and again the RRF-based method appears relatively robust to non-Gaussian BPD profiles, where the Gaussian-based method fails. In this scenario, fusion with AR, does not give a large change in estimates over just using the RRF since the RRF trees are confident in their estimate; there is



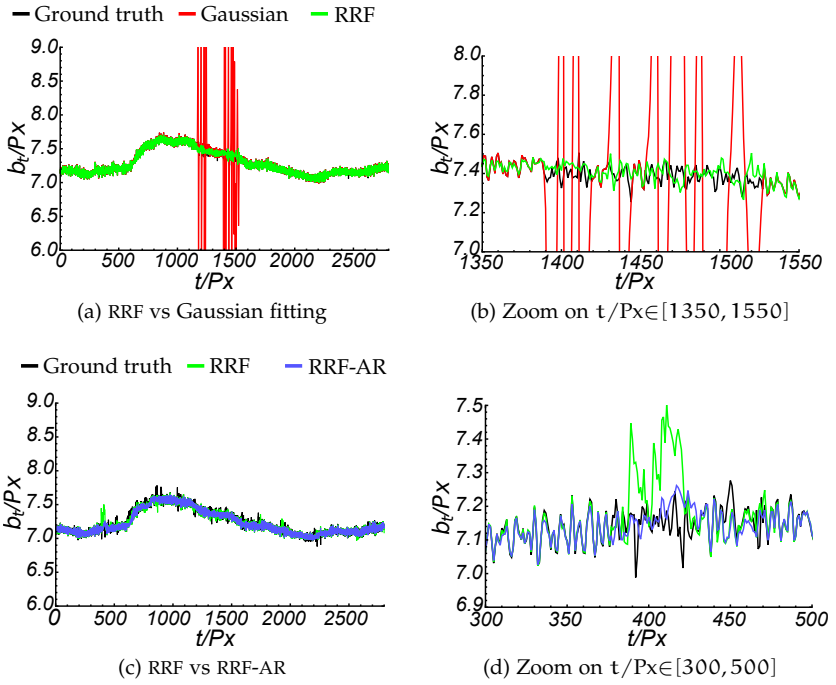
(a) RRF vs Gaussian fitting

(b) Zoom on t/Px∈[300, 500]

(c) RRF vs RRF-AR

(d) Zoom on t/Px∈[300, 500]

Figure 5.11: *Easy* scenario wobble estimates. In (a) the RRF instantaneous estimates (green) are compared to the Gaussian-based method (red), and the ground truth (black). In (c) the RRF-based instantaneous estimates (green) are compared to its Bayesian fusion with an Auto-Regression (blue). Plots (b & d) show zooms for the most difficult region. In (d), the fused estimator yields much more accurate estimates.



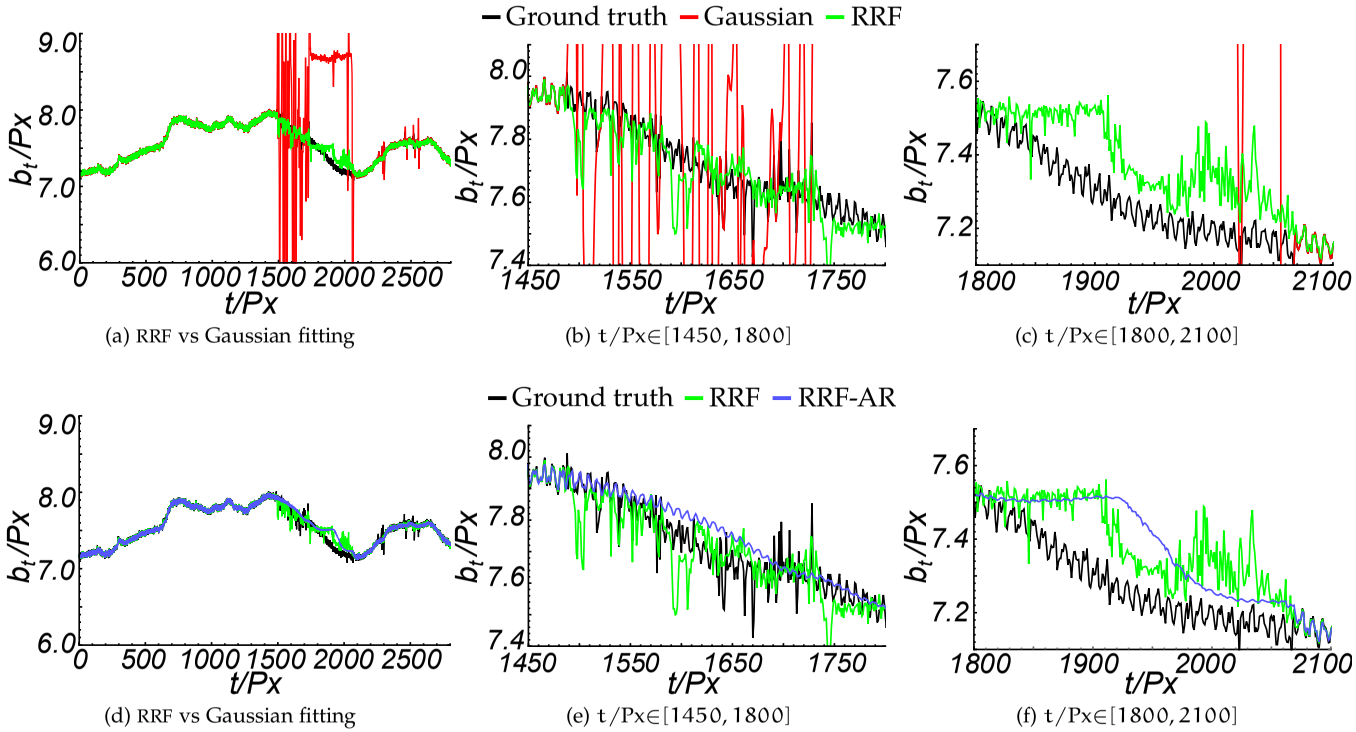
(a) RRF vs Gaussian fitting

(b) Zoom on t/Px∈[1350, 1550]

(c) RRF vs RRF-AR

(d) Zoom on t/Px∈[300, 500]

Figure 5.12: *Intermediate* scenario wobble estimates. In (a) the RRF instantaneous estimates (green) are compared to the Gaussian-based method (red), and the ground truth (black). In (c) the RRF-based instantaneous estimates (green) are compared to its Bayesian fusion with an Auto-Regression (blue). Plots (b & d) show zooms for the most difficult region. In (d), the fused estimator yields much more accurate estimates.



(a) RRF vs Gaussian fitting (b) t/Px∈[1450, 1800] (c) t/Px∈[1800, 2100]

(d) RRF vs Gaussian fitting (e) t/Px∈[1450, 1800] (f) t/Px∈[1800, 2100]

Figure 5.13: Beam position estimates for the *difficult* scenario. In (a) the RRF based method (green) for instantaneous estimation is compared to the Gaussian-based method (red), and the ground truth (black). In (d) the RRF-based method (green) for instantaneous estimation is compared to its Bayesian fusion with an Auto-Regression (RRF-AR; blue), and the ground truth (black). Plots (b, c, e, & f) show zooms for the most difficult regions. The RRF struggles to give accurate estimates in (c), because the BPD is passing across a truck engine, which is very dense and therefore the SNR is very low. This increases the RRF uncertainty, and so the fused estimate puts full weight on the AR estimate, which results in a constant estimate (f) until better RRF estimates are achieved. So the AR has forced the Bayesian fusion into giving sensible estimates. This also occurs in (b) and (e) but to a much lesser extent.

not a large amount of variability in their votes. However, an improvement is seen in Figure 5.12d.

For the *difficult* scenario in Figure 5.13, the Gaussian-based method does even worse. The RRF-based instantaneous estimator appears far more robust, with estimates much closer to the ground truth, however the performance is not as great as in the *easy* and *intermediate* scenarios. The fused estimates exhibit a bias (see Figure 5.13f) where the RRF performs poorly over a long time period. This happens where the total signal on the BPD is close to the background noise level (it is very heavily occluded by a truck engine), and hence the RRF finds it difficult to make accurate estimates of the beam position. This is reflected in the RRF uncertainty, and so the fused estimate puts full weight on the AR estimate, which results in a constant fused beam position estimate until a good instantaneous estimate is achieved. Therefore, the AR has forced the fused estimate into giving sensible estimates. Since the BPD signal is so low in this object and it occupies a large number of time-points, it would be very difficult to obtain an accurate instantaneous estimate by any method based on the employed BPD set-up.

Scenario	Meth.	Acc.	Bias	Prec.	5%	1%	0.1%
Easy	Gauss.	0.105	**0.003**	0.105	0.300	0.818	2.085
	RRF	0.033	0.004	0.033	0.122	0.168	0.203
	RRF-AR	**0.030**	0.006	**0.029**	**0.104**	**0.140**	**0.167**
Intermediate	Gauss.	0.470	-0.010	0.470	1.751	3.523	5.846
	RRF	**0.019**	-0.001	**0.019**	0.073	0.125	0.170
	RRF-AR	0.021	**-0.001**	0.021	**0.069**	**0.111**	**0.149**
Difficult	Gauss.	0.637	-0.135	0.623	2.012	3.329	5.670
	RRF	0.052	**-0.008**	-0.052	**0.188**	0.262	0.310
	RRF-AR	**0.052**	-0.014	**0.050**	0.191	**0.253**	**0.284**

Table 5.1: Performance metrics for: (i) the Gaussian (Gauss.) based instantaneous estimator; (ii) the proposed RRF-based instantaneous estimator; and (iii) the Bayesian fusion of the RRF estimator with an AR (RRF-AR). The metrics computed are: Accuracy (Acc.); Bias; Precision (Prec.); and the MAE for the worst 5%,1% and 0.1% of estimates. RRF gives an order-of-magnitude improvement over Gauss. for easy/intermediate scenarios, and 3-fold for the difficult scenario. RRF-AR gives 3-15% improvement in MAE across scenarios.

For each of the scenarios (*easy*, *intermediate*, and *difficult*), the performance of the methods was quantified in terms of: accuracy; bias; precision; and MAE for the worst 5%, 1%, and 0.1% of time-points. For each scenario, two images (of scissor lifts, fork-lifts, and trucks for easy, intermediate and difficult, respectively) were used for evaluation. Since each image had four corres-

ponding BPDs, and each BPD measured $>2\times10^3$, this amounted to $>1.6\times10^4$ time-points per scenario for evaluation. The worst MAEs are included since particularly bad time-points can be lost in the accuracy, precision, and bias metrics, particularly if there are many air-only time-points where estimation is straightforward. Moreover, wildly inaccurate wobble estimates could lead to column artefacts in the image after correction so are undesirable. The results are given in Table 5.1. All metrics were averages over all time-points for the given scenario.

For the *intermediate* and *difficult* scenarios, the RRF-based instantaneous estimation offers roughly an order-of-magnitude improvement across all metrics, over the Gaussian-based method. For the *easy* scenario, this improvement is approximately 3-fold; the Gaussian method is already quite good at dealing with simple objects. By fusing the RRF with the AR (RRF-AR), the performance increases across most metrics, particularly for worst MAEs, however, there is little change (or a slight worsening for the *intermediate* scenario) in the overall accuracy. In the *easy* scenario there is roughly a 15% improvement in the MAE for the worst 5% of time-points. For the *intermediate* case the improvement drops so about 5%. Finally, for the *difficult* scenario the worst 1% MAE improves by about 3%.

Whilst these results show an improvement on the prior work [11], it is unclear of how well these results generalise to different images, since only a few images were available for the purposes of testing. Future testing should include many more scans of more variable objects.

5.4.3 *Image correction*

The image correction method was first assessed on an air-only scene. For air-only images, wobble estimation is straightforward, since the BPD profile is not distorted by obscuring objects in the scene. However, air-only images allow one to visualise and fully quantify the improvement from wobble correction. The image quality was assessed based on the fact that a perfect (normalised) transmission air image would have all pixel values equalling unity. Image precision can therefore be assessed by computing the Root-Mean-Square (RMS) deviation or Peak Signal-to-Noise Ratio (PSNR) from this ideal.

Figure 5.14 shows air-only images from traverse and portal mode scans and their full correction split into stages. The stages are: sensitivity R_y correction (Figures 5.14b and 5.14f); wobble and ID offset $\exp\left(-(b_{ty} - d_y^2)/2\beta_y^2\right)$ correction (Figures 5.14c and 5.14g); and source variation A_t correction (Figures 5.14d and 5.14h). Images have been intensity clipped (to the same range) so that the wobble effect is visible in Figure 5.14f. Note the visible difference between the images in Figures 5.14b and 5.14f, this difference is mostly due to wobble. The PSNR drops from 109 dB to 77.2 dB, from portal image (Figure 5.14b) to traverse image (Figure 5.14f) due to the wobble artefact. After wobble correction, to obtain Figure 5.14g, most of the wobble artefact is visibly improved. Indeed, the wobble correction improves the PSNR by 21.3 dB but is unable to achieve the portal mode PSNR.

Scan mode	Noise source	Symbol	Before	After	Reduction
portal	*sensor sensitivity*	R_y	0.2305	0.0000	100%
	offset of ID endpoints	$\{\delta_l, \delta_u\}$	0.0013	0.0000	100%
	wobble	b_{ty}	0.0000	0.0000	–
	source variation	A_t	0.0030	0.0000	100%
	photon count	–	0.0029	0.0029	0%
traverse	*sensor sensitivity*	R_y	0.2305	0.0026	99%
	offset of ID endpoints	$\{\delta_l, \delta_u\}$	0.0013	0.0004	72%
	wobble	b_{ty}	0.0185	0.0054	87%
	source variation	A_t	0.0030	0.0004	74%
	photon count	–	0.0029	0.0029	0%

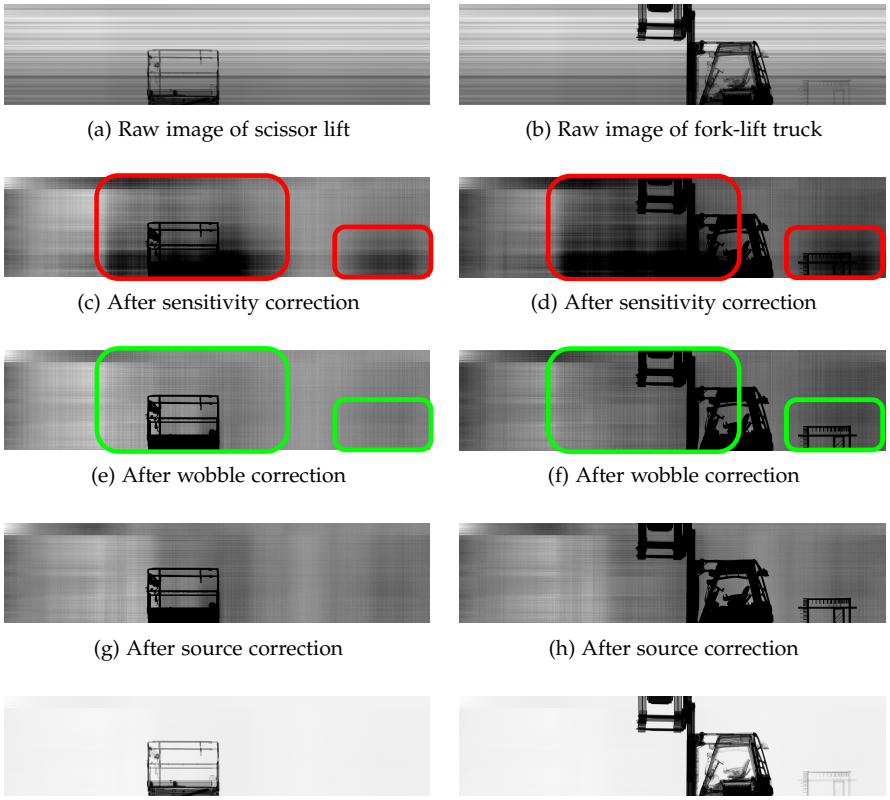
Table 5.2: RMS deviation contributions from different noise sources before and after corrections for: sensor sensitivity; ID offsets; wobble; and source fluctuation. No attempt is made to correct Poisson noise in the photon counts.

Figure 5.14: Correction of portal (a-d) and traverse (e-h) air scans: (a & e) raw images; (b & f) images corrected for for sensor sensitivities; (c & g) images corrected for wobble (and sensor offsets); and (d & h) the image corrected for source variation. Images have been intensity clipped so that the wobble artefact is visible in (b), and the PSNR is given in decibels. The wobble artefact is clearly visible in (f), but not in (b) since portal mode is not effected by wobble. Wobble reduces the PSNR by 31.8 dB. After wobble correction (g) there is a significant visual improvement, and improvement in PSNR of 21.3 dB.

To make quantitative assessment of the effects visible in Figure 5.15, the RMS deviations of the traverse and portal mode air-only images, before and after the different corrections, were used to deduce the magnitude of the noise sources before and after correction. Table 5.2 shows that wobble increases overall image noise, and has also reduced the ability to correct for *sensor sensitivity*, *ID offset*, and *source* variation. Although it is possible to correct for 99% of *sensor sensitivity*, the magnitude of *sensor sensitivity* is so large that it is still the second most dominant source of noise in the corrected image. Source variation was the least successfully corrected and this is apparent in Figures 5.14g and 5.14h, since the corrected images have some slightly visible column artefacts. Finally, it is possible to correct 87% of wobble, thus outperforming the Gaussian-based method [11], which does not incorporate sensor offset estimates into the correction.

The results for corrections applied to traverse mode images of a scissor lift and a fork-lift truck are shown in Figure 5.15. Images have been intensity clipped, to the same range, to make the wobble artefact visible. The wobble correction is obtained using the Bayesian-fused beam position estimate. The red boxes indicate image regions most effected by wobble, and the green boxes show the same regions but after wobble correction. There is a visible improvement in the wobble artefact after wobble correction, showing that a

good level of correction is obtained even when the BPDs pass through dense objects such as a fork-lift truck.



(a) Raw image of scissor lift (b) Raw image of fork-lift truck

(c) After sensitivity correction (d) After sensitivity correction

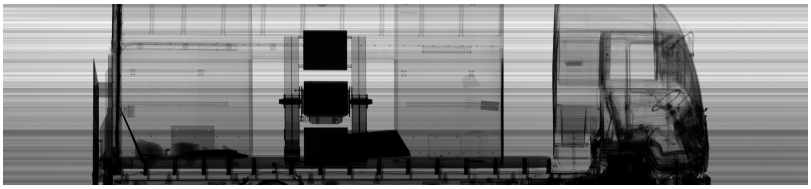
(e) After wobble correction (f) After wobble correction

(g) After source correction (h) After source correction

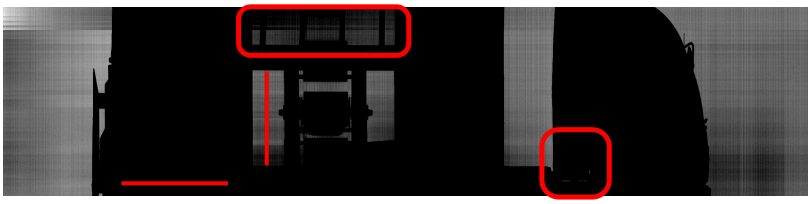
(i) Final image (j) Final image

Figure 5.15: Images of the *easy* scenario (left) and *intermediate* scenario (right), after a series of corrections for: sensor sensitivities (c & d); wobble and ID offsets (e & f); and source variation (g & h). Corrected images were intensity clipped so that the wobble artefact is visible in (c & d). The final images (i & j) are the non-clipped versions of (g & h). The red boxes indicate regions where wobble is visible, and the green boxes indicates these regions after wobble correction. There is a visible improvement in the wobble artefact after correction, thus wobble correction works well even when BPDs are heavily occluded by dense object.

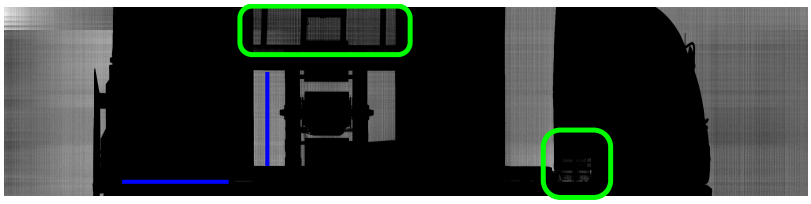
Figure 5.16 shows image corrections on a truck image. Since the truck occupies most of the image, it is more difficult to see the effects of wobble and the corrections. The most obvious places are the steps up to the driver's cabin and the area surrounding the test object. These are indicated by the red boxes in Figure 5.16b. After wobble correction (green boxes in Figure 5.16c), the artefact is reduced so that the driver's steps and the test object become visible. Figures 5.16f and 5.16g show plots of a column and row of pixels, respectively. In each, the red plot is from Figure 5.16b before wobble correction, and the blue plot is from Figure 5.16c after wobble correction. The pixels are taken from image lines that should have approximately constant



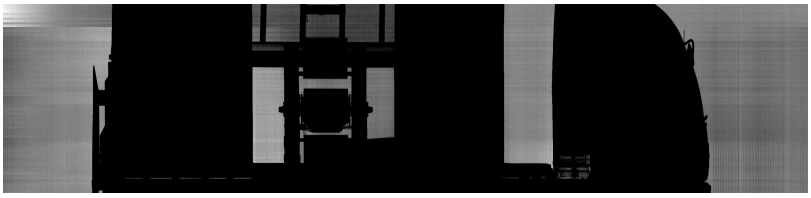
(a) Raw image of a truck



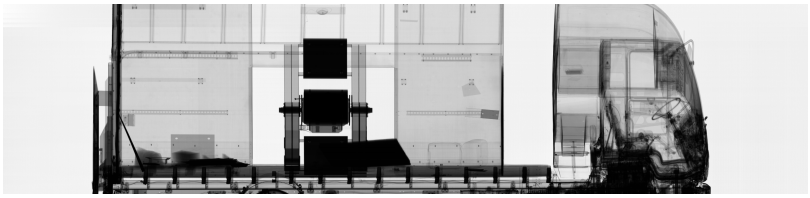
(b) After sensitivity correction



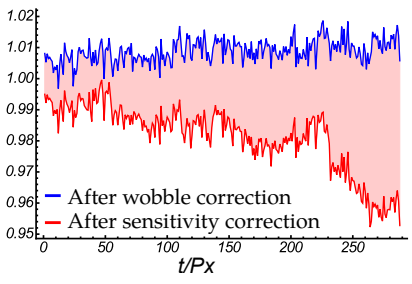
(c) After wobble correction



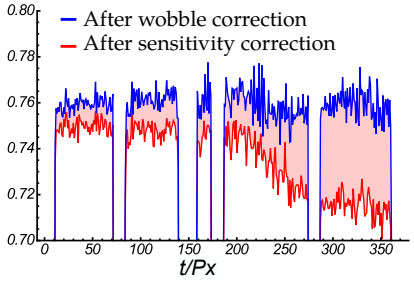
(d) After source correction



(e) Final image



(f) Pixel column traces



(g) Pixel row traces

Figure 5.16: Images from a traverse mode scan of a truck, after a series of corrections. The red boxes indicate regions around the driver's steps and a test object, where wobble is particularly visible. The green boxes indicates the same regions after wobble correction and one can see a visible improvement in the wobble artefact. The plots (f) and (g) show traces of the pixel intensities across a column and row in the image, respectively. The red traces are uncorrected for wobble and taken from (b), whilst the blue traces are corrected for wobble and taken (c). The red traces should be approximately (piece-wise in g) constant, however they are distorted by wobble. The wobble correction corrects most of this distortion.

(or piece-wise constant) pixel values. However, due to the wobble artefacts they are distorted from constancy. The wobble correction corrects a large part of this distortion.

5.5 DISCUSSION

A series of image corrections to ameliorate detector wobble artefacts and noise in large-scale transmission radiography have been proposed. The corrections were derived by considering a model of X-ray image formation in the presence of a wobbling detector. The correction relies on the estimation of a number of fixed system parameters and dynamic parameters which vary during a scan. The fixed parameters include sensor sensitivities, sensor misalignments, and the width of the X-ray fan-beam. The dynamic parameters include the position of the beam at different points along the detector array, and the fluctuation of the number of photons emitted by the source. A method was proposed for estimating the fixed system parameters by model fitting to an air calibration image.

Wobble is more difficult to estimate, and an online approach is adopted using BPDs. BPDs are placed perpendicular to the imaging array, and measure the cross-sectional profile of the photon beam after interaction with the scene, allowing the position of the beam to be determined and hence detector wobble to be measured. In this chapter, an instantaneous estimator based on a RRF was proposed. The true beam profile is first estimated, as if the beam had not been attenuated by the scene, and then the beam position and its uncertainty are estimated by taking the mean and standard deviation of the responses from a RRF, respectively.

To test the wobble estimation and image correction methods, image data collected of several objects ranging in difficulty from a small scissor lift to a large truck, was employed. A commercial scanner, modified by rotating four IDs by 90° to act as BPDs, was employed. The RRF-based approach to instantaneous estimation performs significantly (an order of magnitude in most cases) better than a naive approach based on Gaussian fitting. Moreover, its fusion with an AR achieves results close to ground truth, even for difficult objects, and performs better than the RRF by 3-15% in the worst cases. It struggles when the object has a low SNR for long durations in the scan, and this problem may not be solvable solved by wobble estimation based solely on BPD readings, unless one can accurately predict future beam

positions from a limited number of accurate prior position estimates. This is unlikely due to the stochastic nature of wobble originating from uneven scanning surfaces or wind. Incorporation of measurement devices, such as accelerometers placed along the imaging array, may improve estimates even where there is almost no BPD signal due to object occlusion.

The wobble and system parameter estimates were used to apply corrections to images. Corrections were applied to traverse and portal mode air-only images and resulted in a reduction of 87% of image error due to detector wobble. This extends the work of Rogers et al. [11], which achieved a wobble correction of 70% The wobble correction method was also applied to difficult images of objects and a notable qualitative improvement in the intensity-clipped image quality was observed, clarifying dense regions of the scene. The method should also allow for improved material discrimination in images captured from dual-energy scanners in traverse mode. State-of-the-art material discrimination, for cargo, is performed by taking the log-ratio (or difference) of images at different energies, and relies on subtle differences between the images [5, 102, 142]. But in commercial traverse-mode systems material discrimination is often inaccurate due to image noise, including from wobble (Figure 5.2). And so wobble correction as a pre-processing step could help improve material discrimination accuracy.

6

THREAT IMAGE PROJECTION FOR CARGO

This chapter proposes a method for Threat Image Projection (TIP) in cargo. Rather than using it to improve and evaluate the performance of human operators as is conventional, it is instead proposed use is in the training and testing of Machine Learning (ML) based algorithms, as it provides a means of data augmentation to overcome the *data problem* which is particularly severe in security domains. It also allows manipulation of testing conditions to evaluate the performance of the system in a number of scenarios. The TIP method is validated both qualitatively and quantitatively using images collected in an experiment with a commercial X-ray scanner. In addition an 'Empty Image Projection' (EIP) is proposed to test whether ML-based systems can learn to exploit potential TIP artefacts to give a falsely boosted performance when tested on TIP threat imagery as opposed to real threat imagery.

6.1 MOTIVATION

One major challenge for obtaining high human performance at visual screening tasks, such as Automated Threat Detection (ATD) for X-ray baggage scans, is the rarity of real threats. Studies have shown that humans perform much better in terms of detection and false alarm rates if threat items have high prevalence [131]. This prompted research into TIP techniques, mostly in Cabin Baggage Screening (CBS), whereby threat items are synthetically concealed in baggage imagery to increase threat prevalence during live screening operations. TIP is also used in Computer-Based Training (CBT) [85, 143], and for evaluating operator performance and vigilance [144].

Most TIP methods insert a Fictional Threat Image (FTI) from a threat database into the image [80]. Researchers have focused on determining realistic placement locations (voids) in baggage and generating threat noise and artefacts that are consistent with the rest of the baggage [83, 145, 146], so as to reduce visual cues for operators. To best knowledge, there have been no academic publications on TIP methods for cargo. Authors have commented on

possible cues caused by superposition-based TIP methods for single-view X-ray baggage [146]. In this work a similar superposition approach is followed, but it is demonstrated, experimentally, that it does not lead to any obvious visual cues.

Researchers also face a similar threat prevalence issue when training ML-based ATD algorithms. There is often a large imbalance between the *benign* and *threat* classes. This can lead to learnt algorithms that are biased towards the *benign* class with suboptimal performance on the *threat* class. This observation is similar, and possibly analogous, to the one found in humans. Class imbalance can also affect performance evaluation, particularly simplistic accuracy measures, in what is known as the 'accuracy paradox' [147].

To remedy the class imbalance problem, researchers often consider:

(i) dataset re-sampling [148, 149];

(ii) reformulating the problem as a one-class problem where only benign data is required [150];

(iii) adjusting the algorithm cost function [151]; or

(iv) generating or collecting more data.

Recently, with the development of end-to-end learning methods such as deep learning, which require very large amounts of training data, dataset augmentation [152, 153] has become increasingly the focus of attention. In dataset augmentation, class-preserving transformations are made to existing training data to expose the ML algorithm to natural variation, which reduces overfitting and improves generalisation to unseen examples. Such transformations often include rotations, translations, reflections, and changes in illumination and noise.

Cargo screening is faced with a major class imbalance problem since threats are extremely rare in the wild. It is also expensive and time consuming to collect large numbers of realistic *staged* threat examples. To this end, it would be beneficial to develop a TIP framework for cargo. The framework would allow generation of realistic synthetic threat images and the injection of realistic variation derived from the characteristics of X-ray cargo image formation. These variations include: (i) translations; (ii) rotations; (iii) pixel noise; (iv) magnification; (v) illumination; (vi) volume and density; and (vii) obfuscation. Whilst TIP is beneficial in training ML-based algorithms, it is also useful for gaining a deeper understanding of algorithm performance by controlling particular aspects in testing.

The TIP framework is applied during training and evaluation of *image understanding* algorithms in Chapters 7 and 8.

6.2 THREAT IMAGE EXTRACTION AND PROJECTION

It is assumed that X-ray image formation obeys the Beer-Lambert rule so that the pixel value I_{xy} at image location $\{x, y\}$ is given by

$$I_{xy} = I_0 \exp\left(-\int \mu_{xy}(z)\, dz\right), \tag{6.1}$$

where I_0 is the beam intensity, x are horizontal image coordinates, y are vertical image coordinates, z are depth coordinates, and μ is the attenuation coefficient of the objects composing the scene. The TIP method derived here, assumes no X-ray scatter. Although approximate, this allows rapid synthesis TIP imagery, which is required when augmenting data on-the-fly for training the ATD algorithm in Chapter 8.

The pixel value can be split into contributions from the threat T and its background B, such that

$$\begin{aligned} I_{xy} &= I_0 \exp\left(-\int_T \mu_{xy}(z)\, dz\right) \exp\left(-\int_B \mu_{xy}(z)\, dz\right), \\ &= I_0 T_{xy} B_{xy}, \end{aligned} \tag{6.2}$$

Therefore, by estimating $I_0 B_{xy}$, one can estimate the *threat attenuation* $T_{xy} \in [0, 1]$.

Two methods are proposed for extracting the *threat attenuation*, depending on the difficulty of the extraction. If the original threat is on an empty background, then it is simple to estimate this background by taking a small empty patch above the threat and computing the mean of each of its columns. This is because cargo containers are uniform in appearance in the vertical direction.

This method of extraction is used for generating a TIP database for training an ECV algorithm in Chapter 7.

In more difficult cases, where the original threat is situated on some low-density but non-uniform background, one has to perform a more involved extraction procedure. This includes:

This approach is adopted in Chapter 8, where staged threats require supporting fixtures for imaging.

1. Manually delineating the threat (*threat mask*);
2. Manually delineating background/benign structures (*ignore mask*);
3. Computing the mean pixel intensity of background pixels that are not members of either the threat or ignore mask;

4. Dividing pixels in the threat mask by this mean;

5. Setting all other pixels to unity.

This process gives an estimate of the *threat attenuation*, which can then be projected into X-ray images by multiplication. Unlike TIP for baggage Computed Tomography (CT) [88], one does not have to compute plausible threat locations, unless the threat occupies a very large container volume. An illustration of extraction using masks is shown in Figure 6.1.

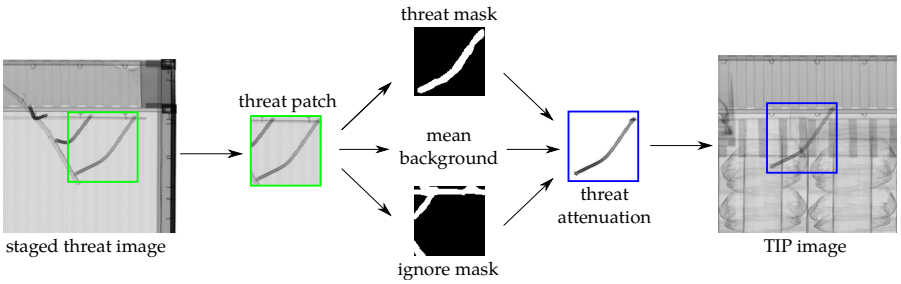


Figure 6.1: Illustration of Threat Image Projection (TIP) when the threat background is complicated. An *ignore mask* and a *threat mask* are manually delineated. The mean is computed for pixels not belonging to either mask. The pixels belonging to the threat mask are divided by this mean, and all other pixels set to unity. The *threat attenuation* can then be projected in new images by pixel-wise multiplication. Original figure from Jaccard et al. [17].

6.3 EXPERIMENTAL VALIDATION

Since the proposed TIP method relies on assuming the Beer-Lambert law and a monochromatic X-ray source, it is unclear how realistic TIP imagery is without experimental validation.

It is important that TIP imagery is realistic, in particular the TIP process should not generate any cues that may be learnt by a ML algorithm, especially if testing is performed on TIP imagery. To this end, the TIP method was validated experimentally, using a Rapiscan® Eagle M60 operating in interlaced dual-energy mode using Bremsstrahlung X-rays with 4 MeV and 6 MeV cut-offs for low and high energy, respectively. Images were captured of containers, containing:

1. *threat only* (T) as shown in Figures 6.2a and 6.2b;

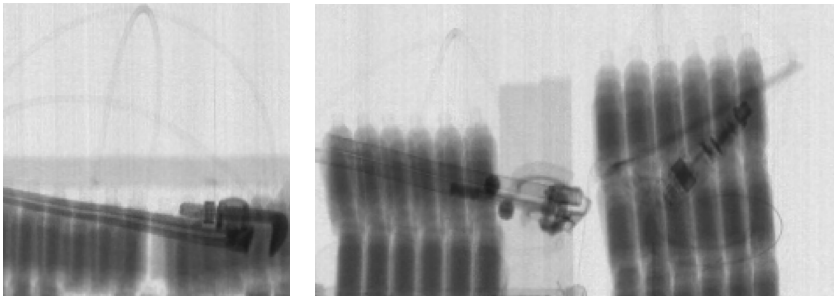
2. *threat and other cargo* (TC) by leaving threats untouched and adding in cargo as shown in Figures 6.2c and 6.2d;

3. *other cargo only* (C) by removing the threats from the container and leaving the background untouched.

Industrial tools (i.e. pipe wrench, electric drill, pipe bender) were used as threat models and plastic cylinders and hula-hoops were used to support the threats in place. For the purposes of this experiment, the hula hoop support structures were included as part of the threat since they were only present in the threat image, and would require more complicated background removal than is proposed here. Moreover, it was impractical to remove the threats from their hula hoop support structures between experiments, so that they could be imaged as background rather than threat.

(a) Wrench (b) Pipe bender and drill

(c) Added background (red) (d) Added background (red)



(e) Captured X-ray (f) Captured X-ray

Figure 6.2: The top row shows photographs of the industrial tools that were used as threat models in this chapter. The second row shows the tools with added background cargo (red). The final row shows the X-ray scans of the threat and background. These X-ray images are compared with the equivalent TIP imagery to assess TIP realism.

Threats were extracted from the T images and projected onto C images to create the TIP image (Figure 6.3). Images were first registered using affine transforms so that the containers were aligned in the image.

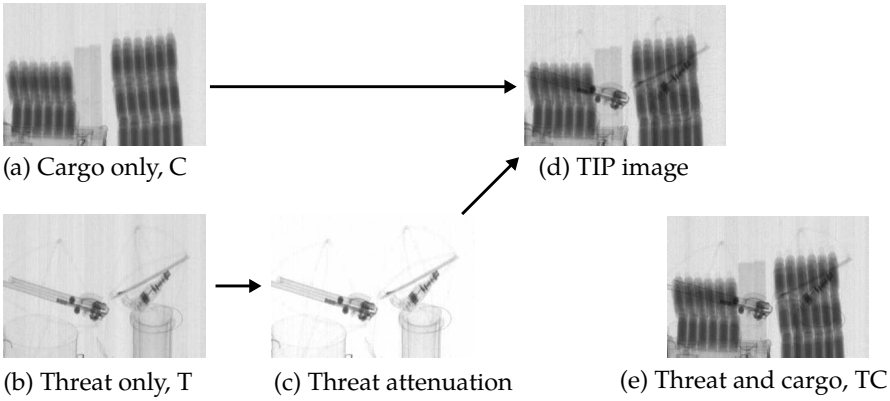


Figure 6.3: Illustration of Threat Image Projection (TIP) used in experimental validation. Three images were captured of: (i) cargo only (C); (ii) threat (and support structures) only (T); and (iii) threat and cargo (TC). The background was removed from T by dividing by a background estimate leaving the *threat attenuation* mask. The threat was then projected onto C by multiplication. For evaluation the TIP image can be compared to the treat threat image TC. Note that the threat support structures have been included as threat in this experiment, but would be removed in practice.

Visual comparison of TIP images and TC images (Figure 6.4) shows that TIP is realistic; one would not be able to distinguish which image is real and which is TIP without being told. Furthermore, the TIP error can be quantified by measuring the deviation between the TIP image and the TC image and compared to the natural variation of the system. The natural variation was estimated by taking the deviation of repeat TC scans. In both cases, one can also study the distribution of deviations in a histogram and compute the Peak Signal-to-Noise Ratio (PSNR). This is shown in Figure 6.5. It was observed that TIP does not give rise to large errors relative to natural image variation (TIP error is less than natural variation in this case), and that the errors, in distribution and spatial arrangement, are very similar to natural variation. In addition, there are no obvious visual cues generated from the TIP process.

6.4 INJECTION OF REALISTIC VARIATION (DATA AUGMENTATION)

Variation can be injected into the threat appearance, using transformations that preserve the class of the threat. These transformations can be derived by considering the nature of X-ray image formation. Here several different

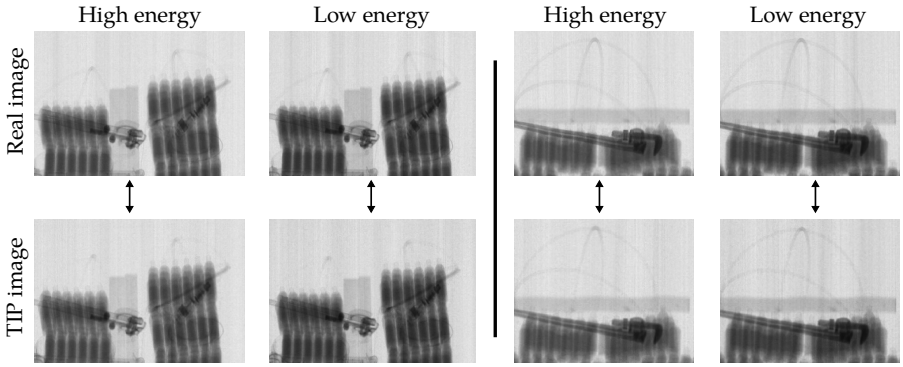
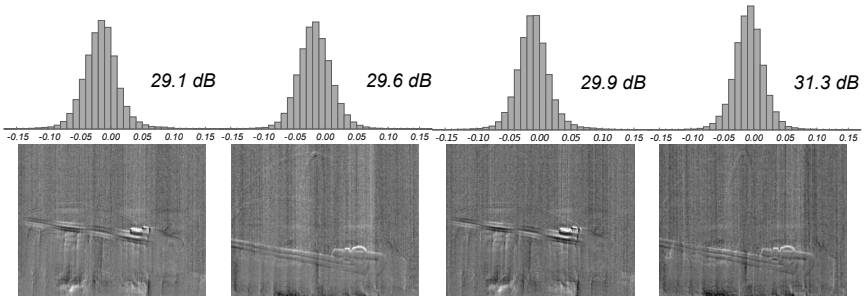


Figure 6.4: Qualitative comparison of real threat images (top) and TIP images (bottom) for high and low energies. Industrial tools (i.e. pipe wrench, electric drill, pipe bender) have been used as a threat model. The images correspond to raw captured data and no additional processing (e.g. denoising) has been applied.



(a) LE: Natural (b) LE: TIP (c) HE: Natural (d) HE: TIP

Figure 6.5: Comparison of natural system variation (a & c) with Threat Image Projection (TIP) errors (b & d) for high energy (HE) and low energy (LE). Natural variation was computed as the deviation between repeat scans of identical cargo and threat (TC). TIP error was computed as the deviation between the TIP image and a TC image. Images show these deviations, rescaled, so that they are visible. Histograms show the distribution of deviations. For each case, the PSNR was computed based on the deviations, and are given in decibels (dB). TIP does not lead to large errors relative to natural image variation and does not change the distribution of deviations.

types of transformations applicable to X-ray cargo imagery are discussed. These are class-preserving transformations and can thus be used for data augmentation.

6.4.1 *Translations*

Translation variation can be injected either by controlling the threat insertion position (Figure 6.6b), or by oversampling the threat item. Oversampling, samples multiple windows that overlap the threat, but with a small displacement. It is similar to random crops, which been used in the wider computer

vision community [152] to encourage learning of small translation invariant features and to achieve class balance. However, TIP also allows one to vary larger scale placement of the threat (e.g. within a cargo container). It is possible to obtain full coverage of possible threat locations within the container.

Whilst TIP enables manipulation of threat placement, it also provides ground truth labels for the threat Region-Of-Interest (ROI) within an image. Such labels are essential for training and testing detection algorithms, and avoids the inconvenience of manually labelling threat ROIs.

6.4.2 *Magnification*

The magnification effect was demonstrated in real X-ray images in Section 3.3.3.

In X-ray cargo scanners, the fan-beam geometry means that photon paths are divergent, rather than parallel, and so the appearance of an object varies as a function of the distance from the source. When the object is close to the source it appears taller in the image than when it is placed further away. Therefore there is a natural variation in the vertical magnification of the object depending on its location. It is proposed that magnification scale can be approximated by

$$\alpha = 1 + d\left(\frac{l_f}{l_n} - 1\right), \tag{6.3}$$

where $d \in [0, 1]$ is the distance away from the source normalised by container depth, l_n and l_f are the vertical lengths (in pixels) of the same object placed at nearest and furthest container wall from the source, respectively. The container walls, themselves, can be used to measure l_n/l_f for a particular system. The parameter d can be sampled randomly when generating TIP examples for algorithm training.

Magnification variation is demonstrated in Figure 6.6a.

6.4.3 *Rotations*

A 3D rotation of a threat, has a corresponding 2D image appearance transformation, which is non-trivial to determine, particularly for out-of-plane rotations. For example, consider a TIP library image of a hand-drill, where the drill piece is oriented toward the source. This appears as a dark rectangle in the image. If one rotated the drill through 90 degrees so that the drill piece is orthogonal to the source, the drill in the image would appear

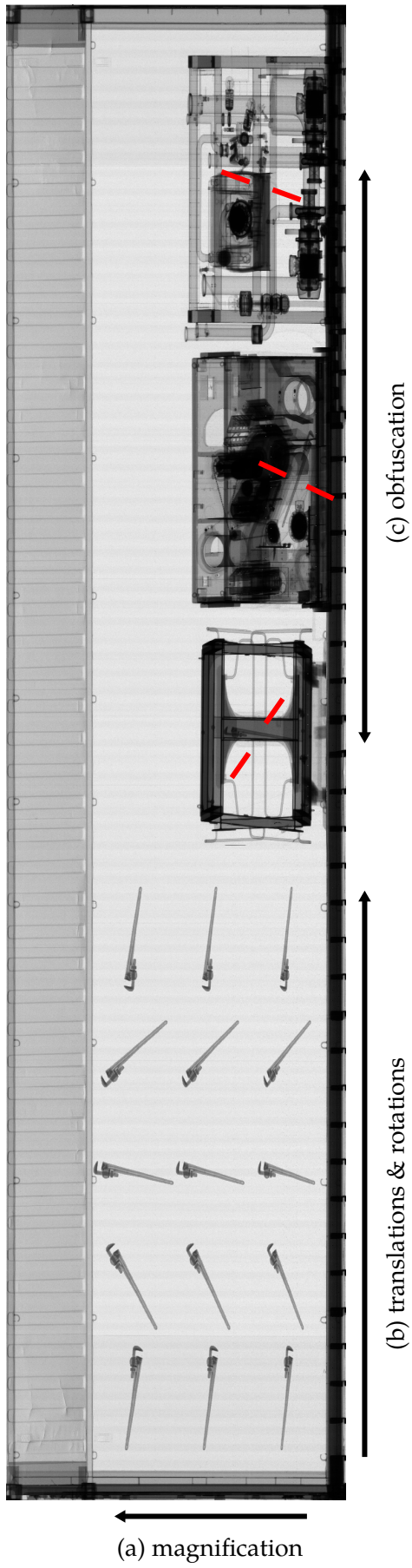


Figure 6.6: Illustration of variation injection for Threat Image Projection (TIP) into an empty cargo container. A pipe wrench has been used as a threat model. On the left of container, and going bottom-top, the vertical dimension of the wrench has been stretched (magnified) to represent appearance variation as it is positioned at varying distances form the source. Going left-right, the wrench is being translated and rotated anti-clockwise. On the right of the container, the wrench has been obfuscated by different database objects before insertion into the container. The red lines indicate the location of the wrench.

L-shaped. Determining a 2D image transformation just from the original image to obtain an L-shape is not trivial, unless one has more information. Such information could include multiple images captured of the drill at different, known, orientations. With this it could be possible to interpolate between 2D images at different orientations to obtain the 2D image transformation. Similarly, if one has 3D information about the drill (such as a CT scan or a 3D model) it might be possible to create realistic 3D rotations.

The TIP libraries used in Chapters 7 and 8, do not have this information. Instead, we only use 2D image rotations, which approximates in-plane rotations. Adding random 2D rotations to threats during training encourages the learning of rotation-invariant features. In case one is concerned that the interpolation used in rotations provides subtle cues, one can apply flips to the threat image instead. However, this restricts the amount of possible variation. Rotations are demonstrated in Figure 6.6b.

6.4.4 *Noise*

Noise types are discussed in Chapter 3.

Cargo X-ray images are mostly affected by: (i) salt-and-pepper noise possibly from bit errors, dead pixels, or analogue-to-digital conversion; and (ii) Poisson noise originating from the number of photons emitted. Both types of noise can be added to TIP imagery for training ML-based algorithms, so that they can become robust to such noise. It is particularly useful to vary the noise on a threat item that is used multiple times in training.

6.4.5 *Illumination*

The illumination (mean number of X-ray photons emitted) can vary for different images due to slight differences between scanners. Illumination can also vary within images due to detector wobble in mobile configurations or due to X-ray source fluctuations [11].

Illumination variation can be injected into training data, by scaling the intensity of an image by some random factor (typically 1 ± 0.05). Variation due to source fluctuation is often removed in an image pre-processing step, but if required can be generated by scaling the intensity of individual image columns by factors sampled randomly from a normal distribution.

This is essentially the reverse process to wobble correction in Chapter 5, but with wobble randomly generated rather than estimated.

Illumination variation due to detector wobble is more difficult to generate, as it varies as a function of image x and y coordinates. However, one could

assume wobble is sinusoidal in x, and determine illumination variation in y by the intersection of the detector array with the Gaussian cross-section of the fan-beam.

6.4.6 *Volume and density*

In some cases, volume transformations leave the threat class intact, for example with bulk powder or liquid narcotics. This is often not the case for threats, where a shrunk sniper rifle does not look like a typical hand-held gun and possibly more like a toy. The relationship between volume and image appearance can be approximated by considering the Beer-Lambert law in Equation (6.1). Scaling the volume V equally in each dimension by some factor ν, i.e.

$$V \to \nu^3 V, \tag{6.4}$$

leads to the transformation

$$T_{xy} \to (T_{x'y'})^{\nu} \tag{6.5}$$

on the *threat attenuation*. The new in-plane coordinate system $\{x'y'\}$ has also been scaled by ν in each dimension. When the volume decreases, the occupied image-area decreases and the threat simultaneously becomes less visible (less attenuating).

In even rarer cases it is useful to scale the density of the threat, such as when detecting container loads as a means of ECV as in Chapter 7. Adding density variations during training makes the algorithm more robust to the possible range of load densities. Scaling the density by p, i.e.

$$\rho \to p\rho, \tag{6.6}$$

approximately transforms the threat as

$$T_{xy} \to (T_{xy})^{p}. \tag{6.7}$$

6.4.7 *Obfuscation*

When smuggling threats, criminals may attempt to obfuscate the threat with benign items to confuse inspectors when performing physical or image-based searches. Obfuscation can include (i) shielding by thick/dense materials so that the threat is barely visible in the image, or (ii) concealing the threat within complex, textured, cargo to make the resultant image very confusing. It is important that ML-based algorithms are exposed to such cases during training, as much as it is for a human. To achieve this one can project threats onto a very diverse range of real Stream-of-Commerce (SoC) images, or can use a database of extracted cargoes to project onto threat items during TIP. The later approach is demonstrated in Figure 6.6c.

Obfuscation is different to occlusion in natural imagery, where opaque objects in the foreground block objects in the background, and information about the occluded object is lost. In cargo imagery, many objects are translucent due to the high beam energies, and so objects tend to be obfuscated rather than occluded.

Controlling the attenuation and complexity of obfuscation can be useful in both training and testing algorithms. For example, it can be ineffective to train an algorithm on threats that are so heavily attenuated that there is almost no information about the threat left. In addition, one might identify that an algorithm is poor at distinguishing threats under certain obfuscation complexities, and may want to encourage the algorithm to perform better in these cases by including more of them in the training data. The mean and variance of the obfuscating attenuation may be suitable measures of difficulty, however this is not investigated in this thesis. Schwaninger et al. [154] have introduced similar metrics in baggage, which may be applicable.

6.5 EMPTY IMAGE PROJECTION

Although every effort has been made to validate, experimentally, that the TIP methodology gives no artefactual cues that can boost the performance of ML-based methods over testing on real data, there is still concerns that subtle artefacts may lead to a small, but misleading, improvement in performance. These subtle artefacts could include halo effects from the threat mask, or slight differences (broadening) in the distribution of noise after projection. The question is: can ML-based systems learn to exploit these subtle cues to boost detection performance? If so, can one train the ML system in such a way that it ignores these artefactual cues?

One way of approaching this problem is to create a TIP process that projects only TIP artefacts and not the threat itself. To achieve this, one can use the same TIP process, but rather than extract threats from images, extract

empty patches. Hence, the name 'Empty Image Projection'. An illustration of Empty Image Projection (EIP) is given in Figure 6.7.

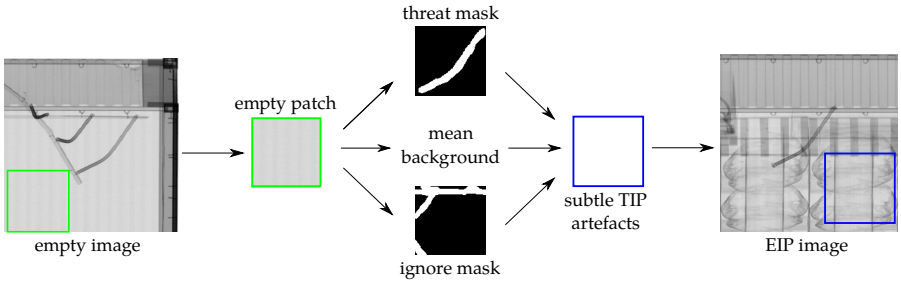


Figure 6.7: Illustration of EIP. EIP is exactly the same as TIP in Figure 6.1, except that it operates on an empty patch rather than a threat patch.

It is proposed that EIP can be used in a number of ways to investigate the effects of TIP artefacts on system performance.

(a) *Feature-level test:* for ML-based systems that operate on hand-crafted image features, one can directly test the features to determine if they are vulnerable to TIP artefacts. For example, if the features are invariant under EIP then there is no information on TIP cues in the features for the ML to learn to exploit.

(b) *ML-level tests:* At the level of the ML algorithm, several tests can be run to verify the extent to which TIP artefacts can be learnt to boost system performance.

(1) *Train on EIP versus benign, test on EIP versus benign.* Train a system on EIP as one class and benign as the other. If the system can learn to detect EIP with greater than chance performance, then it is evident that the cues can be learnt.

(2) *Train on TIP versus benign, test on EIP versus benign.* Test the TIP-trained system on a test set where one class is EIP and the other is benign. If the performance is greater than chance, then it implies that the TIP-trained network has learnt to exploit artefactual cues from the TIP process.

(3) *Train on EIP versus benign, test on TIP versus benign.* Train on EIP as one class and benign as the other, and test on TIP as one class benign as the other. If performance is greater than chance, then it implies that it is possible to obtain some performance just from learning TIP cues.

(4) *Train on TIP versus EIP, test on EIP versus benign*. Use EIP as one class and TIP as the other. The system will learn not to attend to artefactual

cues because they are present in both classes. This can be confirmed if the test yields chance performance.

(5) *Train on EIP versus TIP, test on TIP versus benign*. Same as previous but tested on TIP. This gives an idea of true performance on real data.

Ultimately, the best test for the validity of TIP, is to train a classifier on TIP versus *real threats*. However, a sufficiently large and rich dataset of real threats under different concealments was not available for this thesis.

6.6 DISCUSSION

Training and testing ML-based ATD algorithms is complicated by the difficulty of obtaining large datasets and the major imbalance between the threat and benign classes. The described TIP framework can solve this problem. The framework can be used to generate a very large number of training examples by adding realistic, random, variation during projection, including: (i) translations; (ii) magnification; (iii) rotations; (iv) noise; (v) illumination; (vi) volume and density; and (vii) obfuscation. This framework allows generation of very large numbers of unique training images from a few images captured of a threat, thus enabling rapid addition of detection capability for emerging threats. In addition, it also allows one to form a deeper understanding of algorithm performance by carefully controlling aspects of the test data such as threat position or obfuscation. Both of these approaches are used in the next two chapters.

The threat extraction and projection methods were validated on experimental data and showed no significant qualitative or quantitative difference between TIP imagery and real threat imagery. In particular, there was no evidence that the TIP process created additional visual cues that could be exploited by humans or ML algorithms. Although this is encouraging, there are still concerns that the TIP method produces cues that can be exploited by machine learning based computer vision systems in order to artificially boost performance. As such an EIP framework was proposed to test whether computer vision systems are capable of learning exploiting the cues. In the next two chapters, both TIP and EIP are used in the development of algorithms for ECV and dual-energy ATD.

7

EMPTY CONTAINER VERIFICATION

In this chapter, a method for automated Empty Container Verification (ECV) is proposed. The method operates on full-size cargo imagery and is trained using the Threat Image Projection (TIP) framework proposed in the previous chapter. The main innovative aspect of the proposed system is the use of window coordinates as a feature. This allows the Machine Learning (ML) system to learn the range of appearances at different locations within the container, thus helping to suppress false alarms due to container fabric, damage, and detritus. The system is tested on both real Stream-of-Commerce (SoC) imagery, and TIP imagery representative of small smuggled loads. The Empty Image Projection (EIP) feature-level test also proposed in Chapter 6 is employed to verify that the system is not capable of learning to exploit TIP artefacts as they are not captured by the feature description.

7.1 MOTIVATION

After a completed shipment, containers are either refilled or sent back to the owner empty. As such, a large proportion (20% [155]) of containers in the network are declared-as-empty. Empty containers are often left unsealed, allowing easy access for rip-off attacks.

The amount of contraband smuggled in declared-as-empty containers varies considerably between attempts. Table 7.1 gives examples of foiled cocaine smuggling attempts at the port of Antwerp. Reported seizures of cocaine range from as little as 8 kg hidden in the roof of a declared-as-empty container, to as much as 514 kg hidden within a declared-as-empty refrigerated container [39]. There have also been cases of licit cargo or waste being declared as empty in order to avoid payment of duties [156]. Undeclared non-empty containers can also cause safety problems, for example stacks of 'empty' containers can become unstable and topple over.

Since such large numbers of containers are declared-as-empty, they are rarely, if ever, physically inspected at present. Recent advances in scanner technology have enabled high throughput radiographic imaging of contain-

Throughput is >1 container/s in this work.

Contraband	Origin	Concealment method
59 kg cocaine	Columbia	false partition
109 kg cocaine	Columbia	false partition
4×10^6 cigarettes	Estonia	legitimate partition
514 kg cocaine	Columbia	legitimate partition
8 kg cocaine	Columbia	legitimate partition

Table 7.1: Examples of foiled attempts to smuggle contraband into Antwerp using an empty container or trailer [39].

ers, potentially allowing widespread visual inspection. However, the huge number of images would still make this slow and expensive. Thus, an automated ECV system, capable of inspecting all declared-as-empty containers would be highly beneficial. The system would flag suspicious declared-as-empty images for visual inspection by a human operator.

ECV is non-trivial: (i) the container has inhomogeneous appearance within the image and has parts (such as the corners of the container) that appear similar to loads; (ii) the size and type of the container varies between images; (iii) container damage is visible in the image; (iv) detritus such as garment rails, strapping and empty packaging are often left in empty containers; and (v) loads have a broad range of possible appearances. Figure 7.1 gives some examples of container variation which an ECV system has to cope with without triggering false alarms. In addition, there is a broad variation in the the density, size, texture, and complexity of potential loads (Figure 7.2).

In this chapter, a system for automated ECV in cargo containers is developed. An innovative aspect of this system is the use of window coordinates as a feature, which allows a classifier to implicitly learn the range of appearances at different locations in a container, thus improving classification performance.

7.2 DATA ACQUISITION AND PRE-PROCESSING

In this work the algorithm was trained on realistic synthetic data created using real loads, containers from the SoC, and the TIP framework presented in Chapter 6. Tests were performed on:

(i) *TIP data* – to test the system on examples that are more difficult than typical SoC images, and also to assess the performance of the system as a function of the load difficulty (volume and density); and

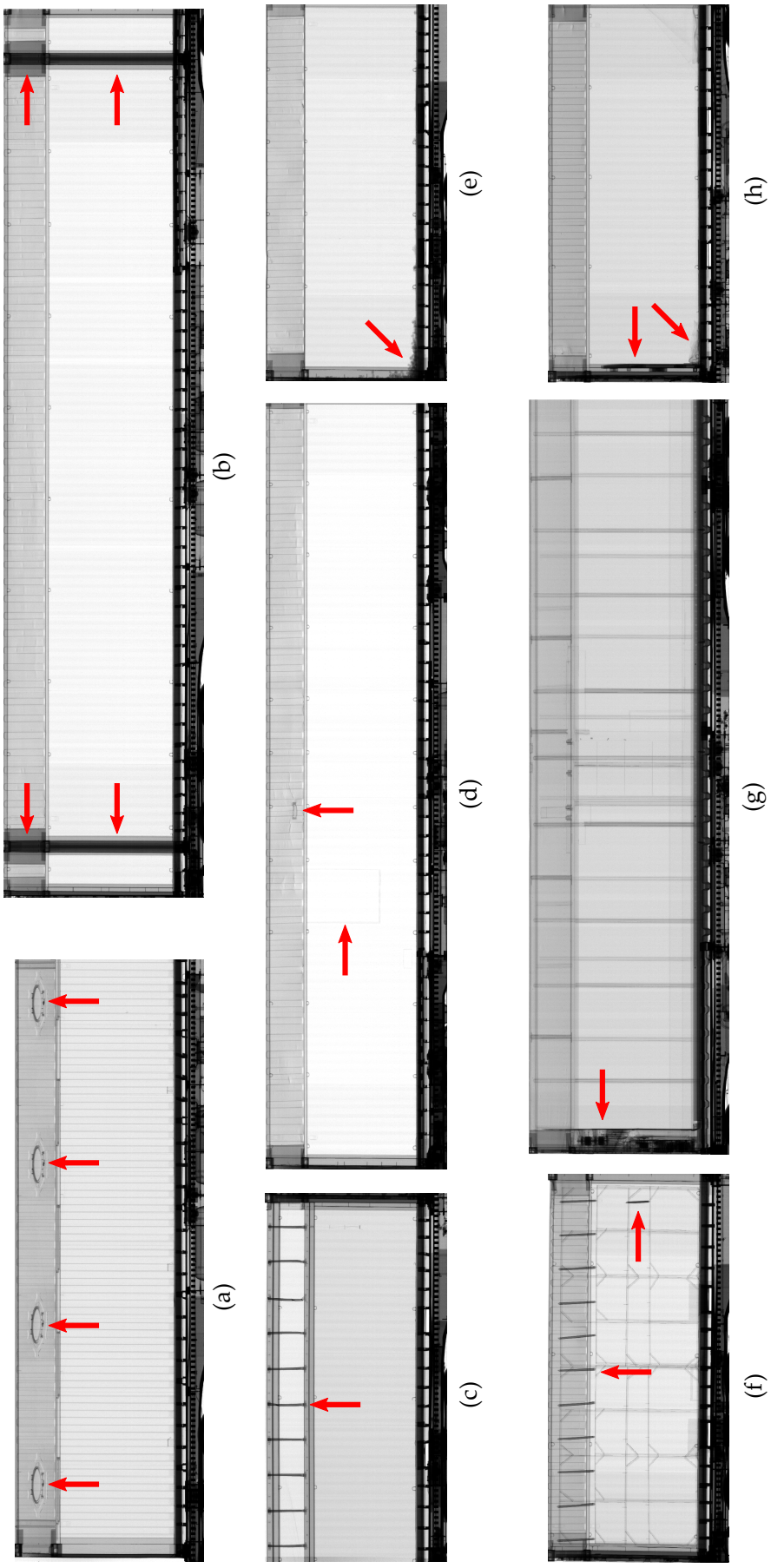


Figure 7.1: Example images of empty containers; (a) a 40 ft bulk container with manholes for inserting dry bulk, (b) a 45 ft High Cube general purpose container with additional posts separated by 40 ft to allow stacking on top of 40 ft containers, (c) a 20 ft open-top container with roof bows to secure a tarpaulin, (d) a 40 ft general purpose container with roof damage and repaired wall damage, (e) a 20 ft general purpose container with detritus, (f) a 20 ft general purpose container with structure for hanging items such as garments, (g) a 40 ft refrigeration unit with insulated walls, and (h) a 20 ft general purpose container with detritus. These examples highlight some of the variation of empty container types that an ECV needs to learn.

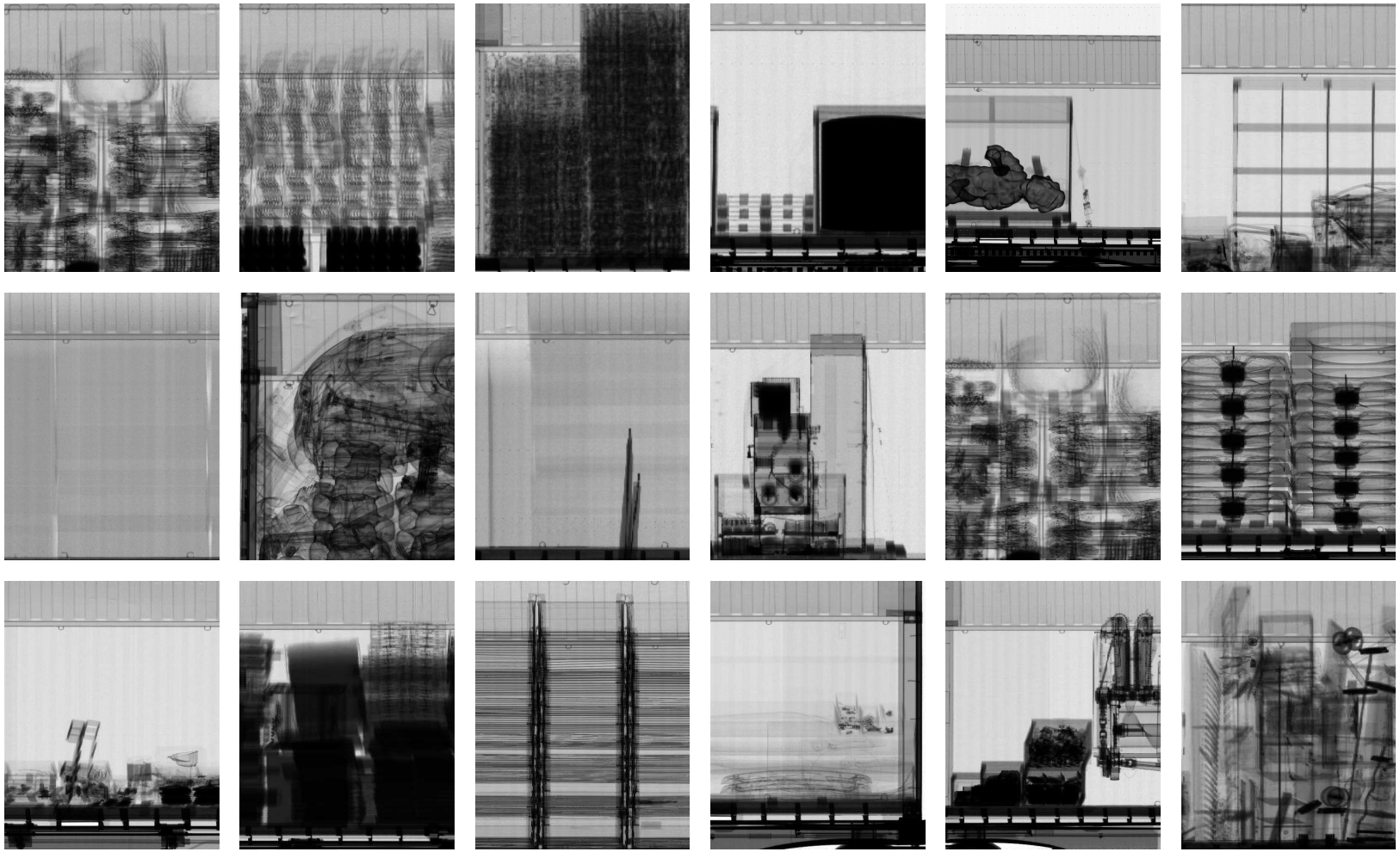


Figure 7.2: Random crops of non-empty cargo containers, demonstrating the broad appearance variation of loads. Loads can range from being very low density with no discernible structure, to very dense with complicated, articulated structure. It is infeasible for an ECV to learn about the detailed appearance of all potential loads.

(ii) *real SoC images* – to check that the system has not overfitted to difficult TIP examples and that the TIP examples are sufficiently realistic.

7.2.1 *Stream-of-commerce data*

The SoC dataset consists of raw dual-energy images from a Rapiscan Eagle® R60. The R60 (Figure 7.3) is a state-of-the-art transmission X-ray system which is capable of scanning containers carried by train, at speeds of up to 60 km/h. Images are warped to correct the effect of any small accelerations of the train as it moves through the scanner. The scanner is fixed and operates in portal mode. The R60 is a dual-energy system which fires interlaced high and low energy beams with 4 and 6 MeV energy cut-offs, respectively. Each image is 16-bit, greyscale, and ranges between 1920×850 and 2570×850 pixels for 20 ft and 40 ft long cargo containers, respectively. The pixel size is 5.6 mm, and the system has effective spatial resolution of the order of a few mm. The image dataset contains a very diverse range of cargoes, including, but not limited to, pallets of commercial cargo, heavy machinery and industrial equipment, household goods, and bulk materials. Approximately 20% of the images are of empty containers.

For ECV only the 6 MeV is used, however in Chapter 8 both the high and low energy images are utilised.

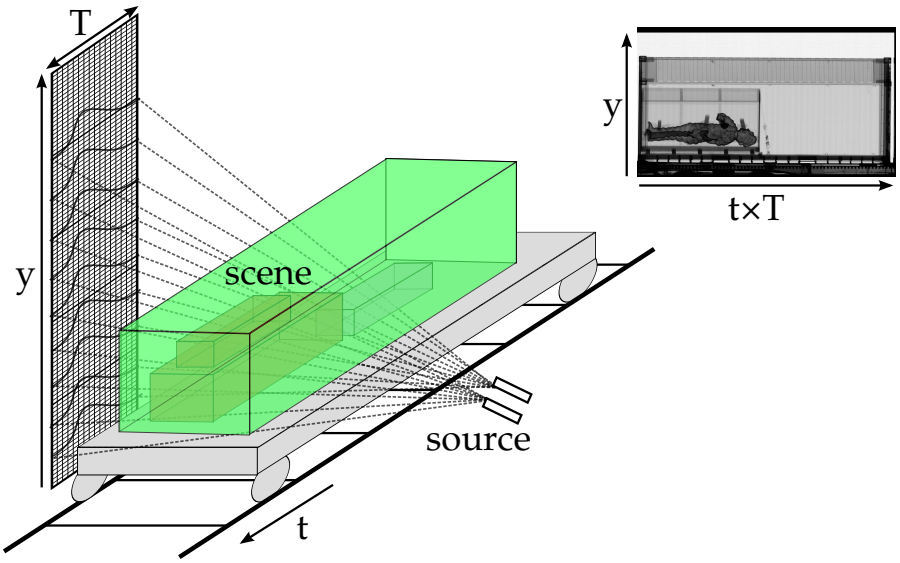


Figure 7.3: Illustration of the geometry of the Rapiscan Eagle® R60 rail scanner. At each time t a small segment of width T=20 is imaged. These are stitched together over time to form the full X-ray image of the container. Source variation [11] leads to vertical stripes (of width T) in the image. The system can scan a full 40 ft container in less than a second.

In total, the SoC dataset consists of 120,000 images. These were manually labelled as *malformed*, *empty* or *non-empty*. Images were labelled as mal-

formed if they contained no practicable cargo information and were excluded from this study. Examples of malformed images are shown in Figure 7.4. Images were labelled as *empty* if they contained no goods; so the *empty* set contains images with other objects such as pallets, packaging and strapping.

Car images were also labelled for use in detection of concealed cars [19, 20], which is not covered in this thesis. Tanker images were removed.

For the SoC testing of the system, 2×10^3 empty and 8×10^3 non-empty SoC examples were collected for the SoC test of the system. The class imbalance reflects the naturally occurring 1:4 ratio between empty and non-empty containers. An additional 6×10^3 SoC images were reserved for the creation of TIP non-empty examples for training and testing. All training and testing sets were kept disjoint, and empty sets used for TIP were kept disjoint from the other empty sets.

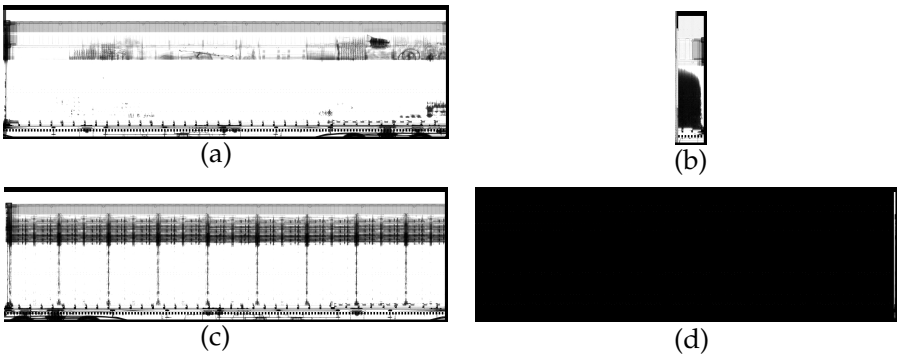


Figure 7.4: Examples of malformed images that were removed from the training and testing datasets. Malformed image contain very little practicable information on the cargo region.

7.2.2 *Image pre-processing*

The raw image had several types of artefact and noise, including: (i) source variation; (ii) sensor sensitivities; (iii) source misfire; (iv) salt-and-pepper noise; and (v) inconsistent container detection. Since the R60 operates in portal mode, no detector wobble occurs. Noise (i) and artefact (ii) are corrected according to Chapter 5. Source misfire leads to black image columns which contain no retrievable image information, and are therefore automatically deleted from the image. To remedy salt-and-pepper noise, isolated dark or bright pixels are replaced by the median of their 3×3 neighbourhood.

The black image columns had zero-valued pixels and occasional salt-and-pepper noise, thus they were straightforward to remove by checking consecutive image columns with the same dimension as the planar imaging array.

The R60 system is configured to automatically fire when a container is entering the scene, and to turn off when no container is in the scene, so that

each container is imaged individually. However, this detection is sometimes inconsistent and a container is scanned as well as a fraction of the previous or next container. Moreover, for the purpose of this work, the only required image parts are those of the container. As such a simple method for segmenting the container region from the image is proposed and tested. The method consists of four steps:

1. *Initial segmentation* – perform a binary segmentation of container (set to one in binary mask) and air (set to zero) using a threshold on image values;

2. *Locate roof* – compute the mode of each row in the binary segmentation, locate the top of the container by searching from the top of image for first image row with a mode of one;

3. *Locate right wall* – search columns to the right of the roof, if there are columns with mode of zero, then set container right wall location as furthest air-column from image edge, else set location at image edge;

4. *Locate left wall* – repeat step 3 for the left hand side of the container.

To test the performance of this simple segmentation algorithm, a total of 480 images were manually segmented to form a ground truth. The automated segmentation was then scored against the ground truth. Since the container is assumed to be rectangular and touching the bottom of the image, the container region can be parametrised by three numbers: the y-position of the roof boundary; and the x-positions of the left and right container boundaries. A poor segmentation was defined as an error in any of the boundary placements of more than 10 Px. This corresponds to $<$6 cm of a 20 ft or 40 ft container. By this definition, two images, were poorly segmented, corresponding to 0.4% of images. These two examples are shown in Figure 7.5.

Figure 7.6 gives examples of the pre-processing applied to a variety of 20 ft and 40 ft containers. The pre-processing successfully deals with a number of difficult cases, including: (i) tanker containers (not used in this work); (ii) multiple containers in the same image; (iii) heavy errors from source variation.

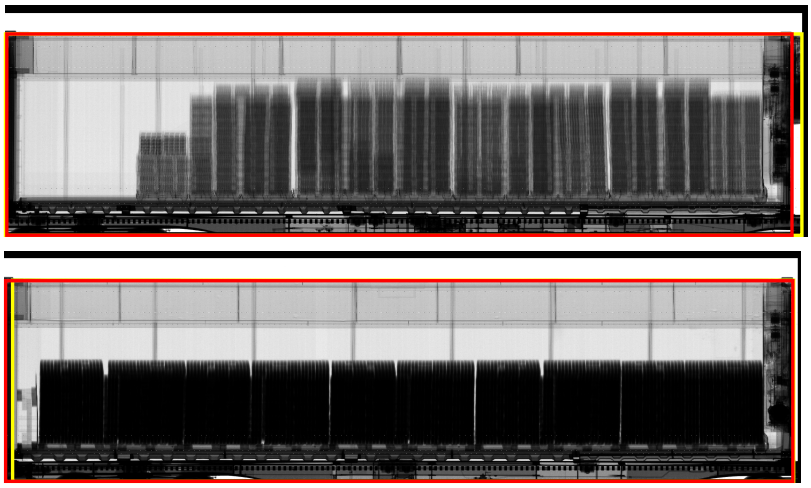


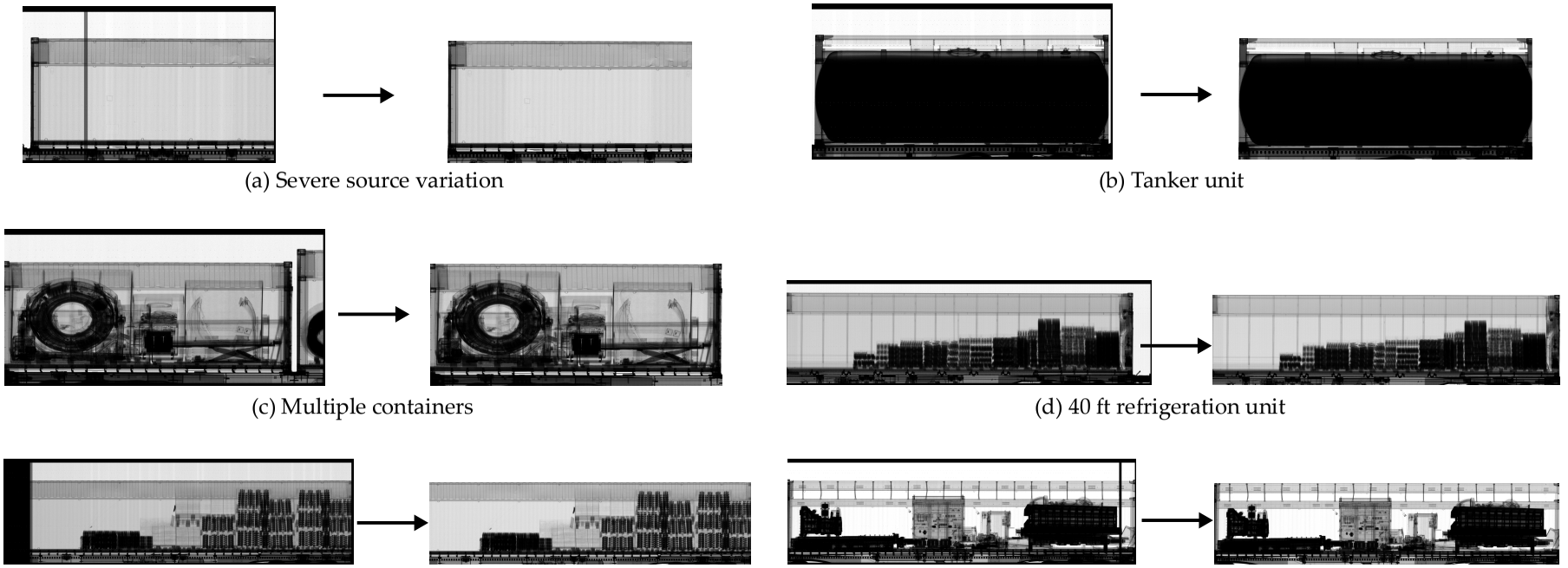
Figure 7.5: The two examples where the container segmentation performed 'poorly'. The red rectangle indicates the manually segmented ground truth, and yellow the automated segmentation. In the top example, the segmentation algorithm gets confused by the external refrigeration unit. In the bottom example, the algorithm erroneously places the segmentation boundary on the inside of the container wall.

7.2.3 *Threat Image Projection*

In the SoC, containers are often packed with goods so that shipping is cost-effective. ECV in these cases is relatively straightforward, since the load occupies a large fraction of the image. In cases where only a small amount of contraband is smuggled, the load is typically much smaller and more difficult to detect. Since there were no difficult non-empty examples in the SoC dataset, they were synthesised from SoC images using the TIP method introduced in Chapter 6. The benefits of the method are that the volume and attenuation of the loads can be controlled which means that (i) the performance can be assessed as a function of load volume and density, and (ii) non-empty examples that are more difficult than the SoC images can be synthesised. TIP also generates a window-level ground truth of the load position, which is convenient for training the system.

A database of SoC loads was created. In this case, loads were simple to extract and no *threat mask* or *ignore mask* was required. The database was partitioned into disjoint databases for training and testing, each with 100 different loads. These loads are projected into SoC empty containers to form realistic non-empty examples. To increase the variation in the appearance of loads, composite loads were formed by:

1. randomly generating a number of loads $1<n<4$;



(a) Severe source variation

(b) Tanker unit

(c) Multiple containers

(d) 40 ft refrigeration unit

(e) Severe source misfire and source variation

(f) 40 ft open-top container

Figure 7.6: Examples of cargo container pre-processing. The pre-processing successfully deals with a number of difficult cases, including: (a) heavy striping from source variation; (b) tanker units; (c) multiple containers in an image; (d) refrigeration units with large air gap; (e) severe source misfire and variation; and (f) open top container with source misfire.

2. selecting n loads at random from the test or train database;

No vertical flipping was included as for large loads this would look unnatural.

3. for each, choose with probability $1/2$ whether to flip the load from left to right;

4. perform a pixel-wise multiplication to form a composite load;

5. transform the composite load to some target volume and density as discussed in Chapter 6.

This procedure is demonstrated in Figure 7.7. By generating a test set with multiple load volumes and densities, the performance can be assessed as a function of load volume and density.

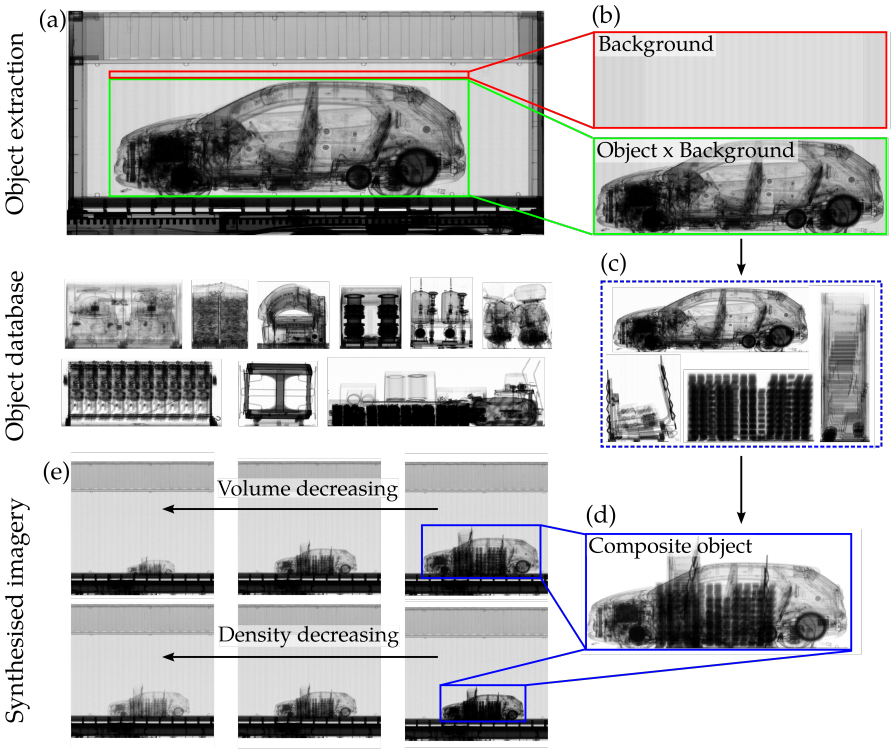


Figure 7.7: Illustration of image synthesis by TIP, performed by; (a) cropping a load (green) and a small background patch (red) out of a non-empty SoC example, (b) performing a pixel-wise division between the cropped load and background leaving an attenuation mask for the load, (c) steps (a) and (b) are repeated to form a database of load attenuation masks, (d) loads are selected at random (dashed blue) and are given random offsets and left-right flips, before pixel-wise multiplication to form a composite load (solid blue), (e) manipulating the volume and density of the composite load before projecting into an empty SoC by pixel-wise multiplication.

Since the log-attenuation (or the probability that a photon interacts in a small volume) is directly proportional to the load density, one can express the load volume and density in SI units by measuring the mean pixel log-attenuation of a load with a known volume and density. Cars were used

for this purpose, since a car is rarely occluded by other loads as it occupies most of the container, making it easier to make accurate measurements. The calculations were based on a typical car having a volume of $12\,m^3$ and a mass of 1.3 t.

When training the system, described in Section 7.3, *non-empty* windows were sampled from the TIP images. Windows were defined as *non-empty* if and only if they have 25% of their area occupied by load or contain at least 25% of the load area. In this way, it was ensured that the training windows 'see' enough of the load, and that the system of tiled windows will always have at least one *non-empty* window whatever the size of the windows and loads.

7.3 PROPOSED ECV SYSTEM

In this section the proposed system for ECV is described in terms of the local window-wise feature encodings, their classification, and their aggregation to determine a classification for the whole image.

7.3.1 *Feature encoding*

Each image was split into a grid of $d \times d$ pixel windows. Windows were placed from the top-left of the image. If an integer number of windows did not perfectly cover the whole image, extra (overlapping) windows were placed tight against the bottom and right edges of the image. For each window a *local feature descriptor* was computed. Experiments were also performed with overlapping windows, where three additional grids (anti-phase in y, anti-phase in x, anti-phase in both x and y) were employed, but there was no noticeable improvement in performance.

A number of local feature descriptors were tested, including image intensity histograms, intensity moments, histograms of Basic Image Features (BIFs), and histograms of oriented Basic Image Features (oBIFs) [157]. Moreover, experiments were performed on using different window coordinates as a feature: including absolute coordinates, normalised coordinates and coordinates measured from the nearest container end. The best performing combination of features was (i) window x-coordinate measured from the nearest container end, and window y-coordinate measured from the top of the container, (ii) oBIF histograms, and (iii) intensity moments. The inferior results

Image intensity histograms are similar to DH descriptors used in baggage [119, 121].

for BIFs, image intensity histograms or different coordinate systems are not presented here. An illustrative overview of the feature encoding scheme is shown in Figure 7.8. Typical performance of BIF driven approaches was 10% lower than oBIFs in terms of detection rate.

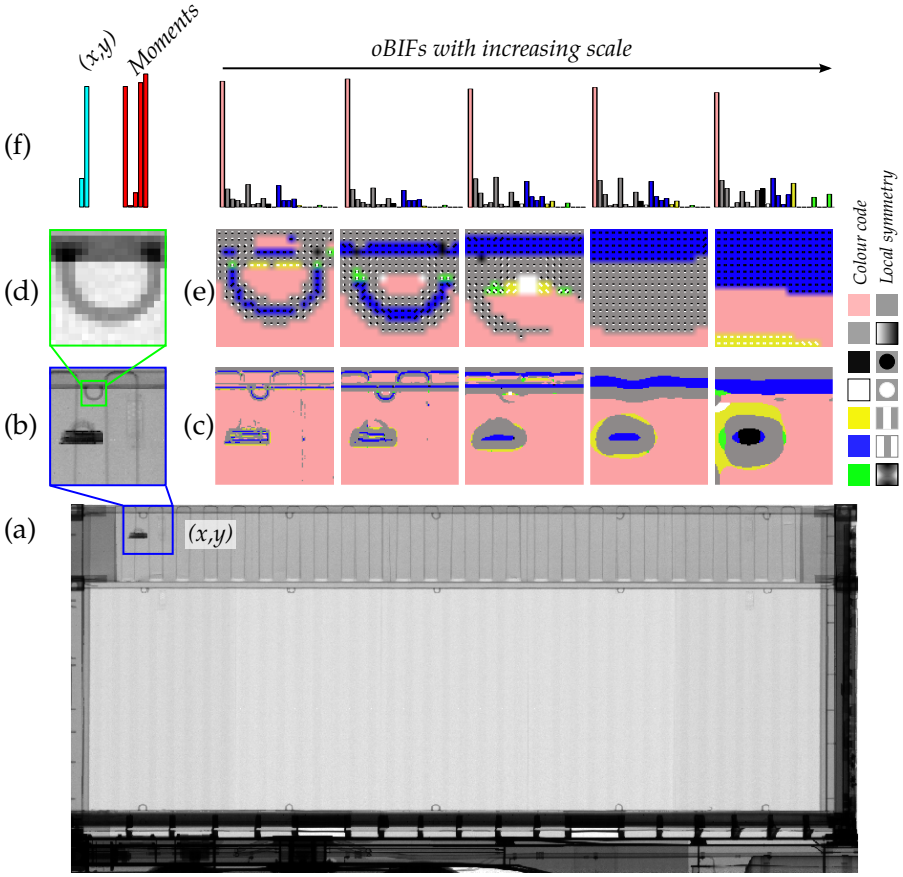


Figure 7.8: Example feature encoding for a 96×96 pixel window (blue) containing a small load. Panels show: (a) the (x, y) location of the window within the image, (b) a zoom of the window, (c) the corresponding BIF primal sketches at different scales, (d) a further zoom of part of the window, (e) BIF primal sketches at this level of zoom with quivers representing the quantised orientation (oBIFs), and (f) the encoded feature vector comprising of the window (x, y) coordinates, intensity moments, and histograms of oBIFs at different scales.

7.3.1.1 *Window coordinates*

By using the window (x, y) coordinates as a feature, the classifier is able, implicitly, to learn the range of background appearances at different locations in the container. For example, it might learn that background near the top of the image should look like roof and that background near the bottom of the image looks like floor, without explicitly being told where the roof and floor are. This is particularly useful in this work since local parts of the image background (container) may appear similar to load. For instance, the

corners of the container appear similar to small dark rectangular loads, but by including window coordinates the classifier is able to learn that small dark rectangular signatures in those locations are likely to be part of the container rather than a load.

It was found that the best performing coordinate system was to measure x from the nearest container end. This reflects the container symmetry, namely that the left-hand side of an empty container looks similar to a left-right flipped version of the right-hand side. This effectively halves the number of x-locations that the classifier needs to learn. The y-coordinate was measured from the top of the container, although it could equally be measured from the bottom.

Using window coordinates as features improves over two plausible alternative methods of learning the container background. One method would be to learn a separate classifier for each (quantised) location. This would require a large number of classifiers to be constructed and stored, and each would be trained on only a fraction of the assembled data. The classifiers learnt from similar areas of the container (e.g. the roof regions away from the ends) would not get to benefit from the training data for the other locations. A second alternative method, which *would* address the training data sharing issue, would be to segment the image into a system of location types (e.g. roof midpoint, roof corner, container centre) and to learn a classifier for each type. However, one would expect the success of this approach to be influenced by the choice of location types and determining these optimally is not simple. In this work, it is proposed that using window coordinates together with Random Forests (RFs), which are capable of expressing highly non-linear decision boundaries, implicitly selects an effective system of location types and groups training data accordingly.

7.3.1.2 *Oriented Basic Image Features*

BIFs [157] are based on responses from a bank of Derivative of Gaussian (DtG) filters up to second order. The vector of DtG responses at a point in the image is known as a local jet. BIFs are derived by partitioning jet space into seven intrinsic regions based on the local symmetry type of the image [159].

Implementations of BIFs and oBIFs in MATLAB are available online [158].

The seven BIF types correspond roughly to; flat, slope, minima, maxima, dark line, light line, and saddle. There are two parameters that control the computation of BIFs; σ and γ. The σ parameter corresponds to the scale of the DtG filter bank used. The γ parameter is a threshold which determines

the amount of structure tolerated before a region is no longer considered flat. Example BIF primal sketches for an image window are shown in Figure 7.8c.

oBIFs are BIF augmented by their local orientation. The flat, minima and maxima BIF types have no intrinsic orientation; slope-like BIFs are augmented by quantising the gradient into one of eight polarised directions; and dark/light lines and saddles are augmented by quantising the second order structure into one of four unpolarised orientations. With the added orientations there are a total of 23 different oBIF types. Example oBIFs for an image window are shown in Figure 7.8e, the quivers indicate orientation.

BIFs and oBIFs have been used for a number of different tasks including; texture classification [160–163], recognition [20, 164–167], and segmentation [18, 168]. In this work, a local distributioon of oBIFs descriptor is formed by computing oBIFs on a given image and within each window forming a 23-bin histogram. Each bin corresponds to a different oBIF type. In this work, integral histograms [169] were employed to compute window histograms efficiently.

7.3.1.3 *Intensity moments*

Moments of the intensity distribution across the window were also computed to encode some information about pixel intensity and its spatial distribution. This information is lost in the locally orderless oBIF histogram, but is useful for the detection of loads particularly for dark loads that have little texture.

For a given image window patch $W_{x,y}$, the intensity moments are defined as

$$\begin{aligned} M_{0,0} &= \sum_x \sum_y W_{x,y} \\ M_{1,0} &= \frac{1}{M_{0,0}} \sum_x \sum_y x W_{x,y} \\ M_{0,1} &= \frac{1}{M_{0,0}} \sum_x \sum_y y W_{x,y} \\ M_{2,0} &= \frac{1}{M_{0,0}} \sum_x \sum_y (x - M_{1,0})^2 W_{x,y} \\ M_{0,2} &= \frac{1}{M_{0,0}} \sum_x \sum_y (y - M_{0,1})^2 W_{x,y}. \end{aligned} \tag{7.1}$$

The five integral image were: $W_{x,y}$, $xW_{x,y}$, $yW_{x,y}$, $x^2W_{x,y}$, *and* $yW_{x,y}$.

The coordinates x, y are taken relative to the window centre.

To compute the intensity moments efficiently for different image windows, five integral images [169] were computed.

7.3.2 *Image classification*

The system uses a RF classifier [170] to score each window. The score for a given window is the fraction of trees that vote that the window contains load. The image score is determined by taking the maximum of the window scores in the image. A threshold is applied to the image score to determine the classification of the image. By varying this threshold, a Receiver Operating Characteristic (ROC) curve can be determined.

In this work, $\sqrt{N_f}$, where N_f is the dimensionality of the feature vector, variables at each split were sampled at random. The number of trees N_T was selected by finding the point where the out-of-bag error (versus number of trees) plot plateaued.

The `randomforest-matlab` implementation of RF [140] was employed.

7.4 RESULTS

In this section the selection of the optimal window size and other system parameters is discussed, next an attempt is made to build a picture of how the system operates, and finally, a detailed analysis is presented of the performance of the system described in Section 7.3 when tested on the SoC and difficult TIP examples.

7.4.1 *Parameter selection*

For oBIFs, the parameter settings γ=0.01 and σ=$\{0.7, 1.4, 2.8, 5.6, 11.2\}$ were used. It was found that using multiple γ did not improve performance, but using multiple scales did. For each window the oBIF histograms, intensity moments, and the window coordinates were concatenated to form a N_f=122 dimensional feature vector. It was found that N_T=512 was sufficient.

To determine a roughly optimal window size, a system, which uses window coordinates and intensity moments as features, was trained and tested at a range of different windows sizes d=$\{32, 64, 96, 128, 160, 192, 224, 356\}$. It was not feasible to use oBIFs as a feature due to the additional computation time required to perform the analysis for each grid size, but this optimum window size selection should generalise (approximately) to the system with oBIFs added. Note that the results of the oBIF system could possibly be improved by repeating this analysis with oBIFs included.

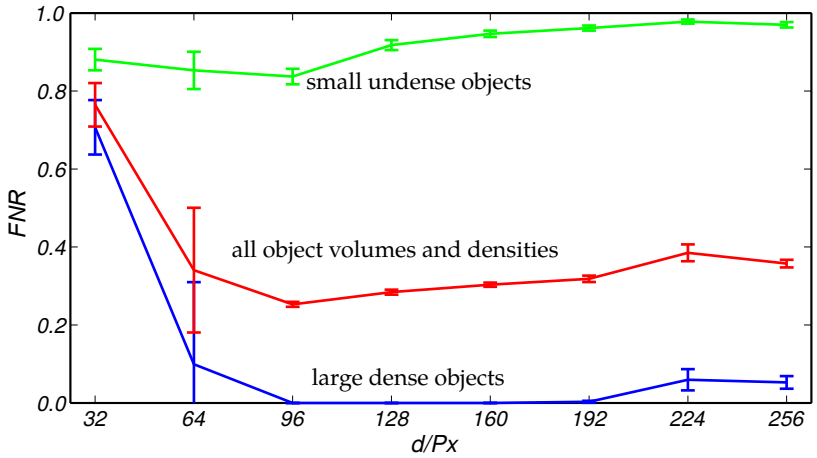


Figure 7.9: The FNR (for fixed FPR of 1%) as a function of window size (d) for easy (large and dense) loads (blue) and difficult (small and low density) loads (green), using intensity moments and window (x, y) coordinates as features. The mean FNR over a range of load difficulties is given in red. Error bars show one standard deviation. The optimum performance occurs at a grid size d=96 Px.

Figure 7.9 shows the average performance as a function of grid size for (i) difficult (small and low density) loads (green), (ii) easy (large and dense) loads (blue), and (iii) a range of difficulties of load. The performance is taken as the average FNR given a fixed FPR of 1%. The average and error bars are obtained by 10 independent trains of the RF classifier. In all cases the best performance is achieved when a window size of d=96 Px is used. There is a performance hit for very small or large windows: if the window is too large, in some cases the load does not excite the features enough; and if too small then windows containing load may appear similar to local parts of the container.

For the rest of the results presented in this section a window size of $d = 96$ Px was employed.

7.4.2 *Empty Image Projection*

The feature-level EIP test proposed in Chapter 6 was employed to test whether potential TIP artefacts are captured in the feature-encoding. EIP was used instead of TIP and feature encodings were computed. These were then compared to feature encodings of the empty container patch without TIP. It was found that the oBIFs were invariant to EIP for the γ and σ settings used, and that the image moments gave a slight difference between empty and EIP patches. This difference was smaller than a pixel and so is unlikely to provide any cues for the ML system to learn. However, to remove any doubt,

the image moments can be rounded to the nearest pixel. An example, showing the invariance of oBIFs to EIP is given in Figure 7.10. The system is also tested on SoC images which adds further evidence that the system has not overfitted to potential TIP artefacts.

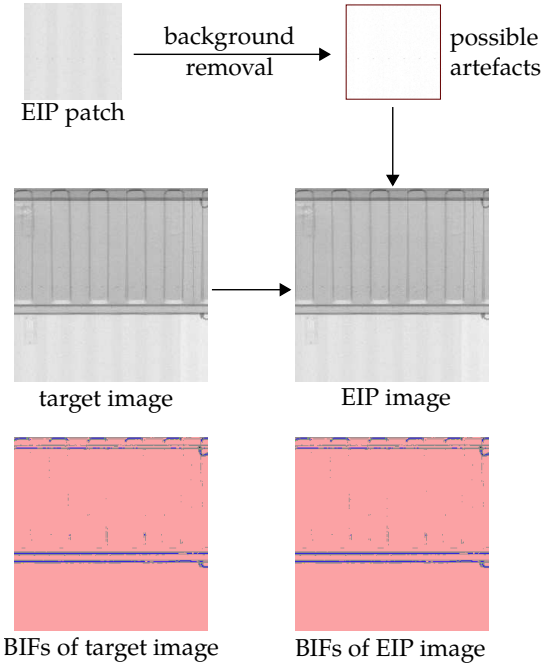


Figure 7.10: An example test of the invariance of the BIF features to EIP. An empty patch is projected using the TIP framework into the target empty container. BIFs were computed on the target empty container before and after EIP and yielded the same result indicating that BIFs, and thus oBIFs, do not capture TIP noise artefacts.

7.4.3 *Testing on the stream-of-commerce*

The system was tested on SoC data. It achieves state-of-the-art Detection Rate (DR) of 99.3% given a 1% FPR, outperforming Orphan et al. [66], who state an accuracy of 97.2% and FNR of 0.4%. This performance implies that the classifier has not significantly overfitted to the TIP examples, and that the TIP examples are sufficiently realistic.

Examples of SoC non-empty detections are shown in Figure 7.11. The blue boxes indicate where a window has been classified as containing a load. In most cases the algorithm is successful at detecting individual load-containing windows, and there are no false positive windows within these true positive image examples.

Score localisation heatmaps of the RF classifier scores were produced. The benefits of heat maps are two-fold. Operationally, they can help an operator

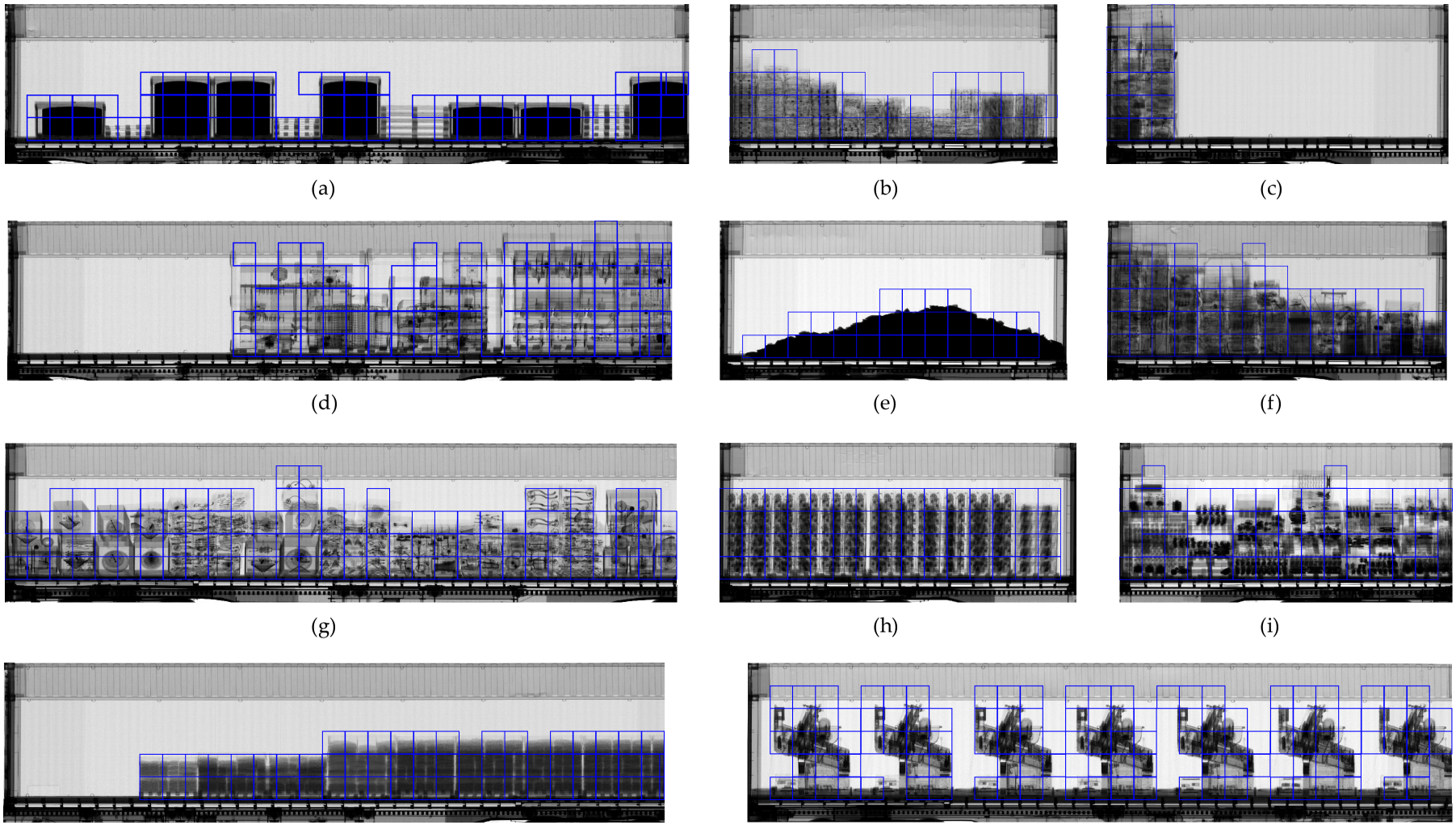


Figure 7.11: Example true positives for Stream-of-Commerce (SoC) images. Blue boxes indicate windows where the score is high enough to trigger a detection.

to focus on the image regions most likely to contain an adversarial load, thus improving inspection speed. For this work, they can also help one understand how the system performs according to the local image appearance. To construct heatmaps, a window was moved across and down the image in steps of 16 Px. At each step the RF score was computed. For each pixel the heatmap pixel value was obtained by averaging the scores of the windows to which the pixel was a member of. The average scores were then rescaled according to the detection threshold. A value of zero implies that the classifier could not distinguish if the window was empty or not, and values of $+1$ and -1 that the classifier was certain that the window is non-empty and empty, respectively.

Figure 7.12 shows score localisation heatmaps for both true positives and true negatives. For SoC loads the detections are well localised to the loads, and the classifier gives a high score for dense loads. Less dense loads achieve a lower, but still positive score. Note that the pallets in Figures 7.12a and 7.12b are given relatively low scores. Pallets should not be classified as loads as there are examples of empty containers that contain left-over pallets. For the true negatives in Figures 7.12g and 7.12h, the classifier yields a low score for all parts of the container. These two examples are relatively straightforward and similar examples are common in the SoC dataset. Figures 7.12i and 7.12j are more difficult because they contain garment rails which locally appear similar to thin loads and are also relatively rare in the SoC. Nevertheless they are correctly classified as empty, but there is some signal where the RF trees are not all in consensus.

Score localisation heatmaps are also given for false positives and false negatives in Figure 7.13. The false negatives are loads that occupy the whole container, and have low density and little texture, and therefore do not excite the features enough to be detected. There are some small excitations where there are thin gaps in the loads. Note that the vertical spikes protruding from the floor in Figure 7.13b are fork-lift pockets and not loads. The false positives are all relatively rare examples in the SoC, particularly the detritus in Figure 7.13c. It is unclear what has caused the darkening of the roof in Figure 7.13g. The false positives on the port holes in Figure 7.13c and refrigeration unit in Figure 7.13d may be improved by augmenting the training data with similar examples.

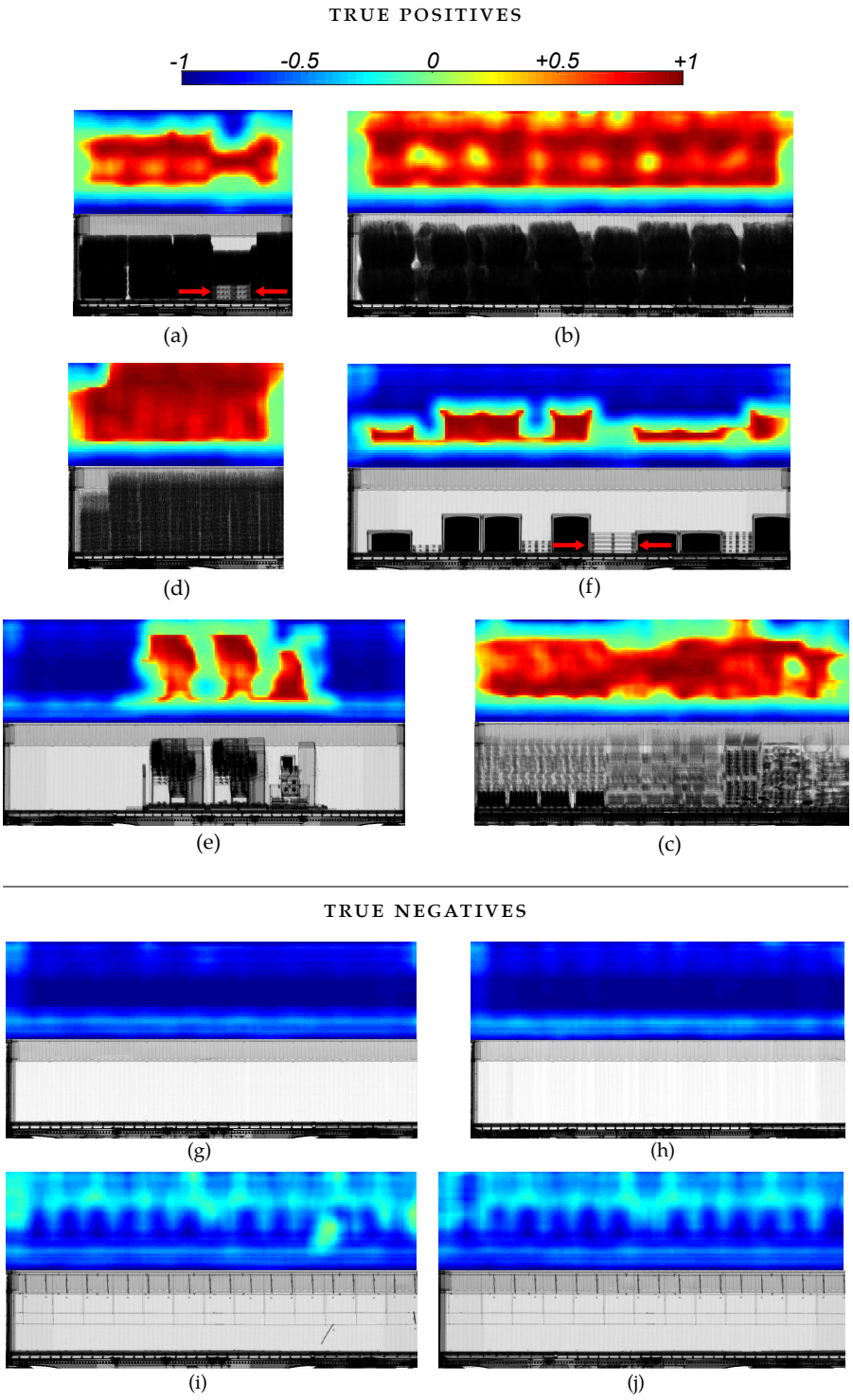


Figure 7.12: Examples of detected SoC loads and their corresponding score localisation heatmaps. Note that the pallets, indicated by the red arrows, in (a) and (f) have been ignored since they can be found in empty containers. In (i) and (j) the roof bows lead to a small detection signal but not enough to trigger a false alarm.

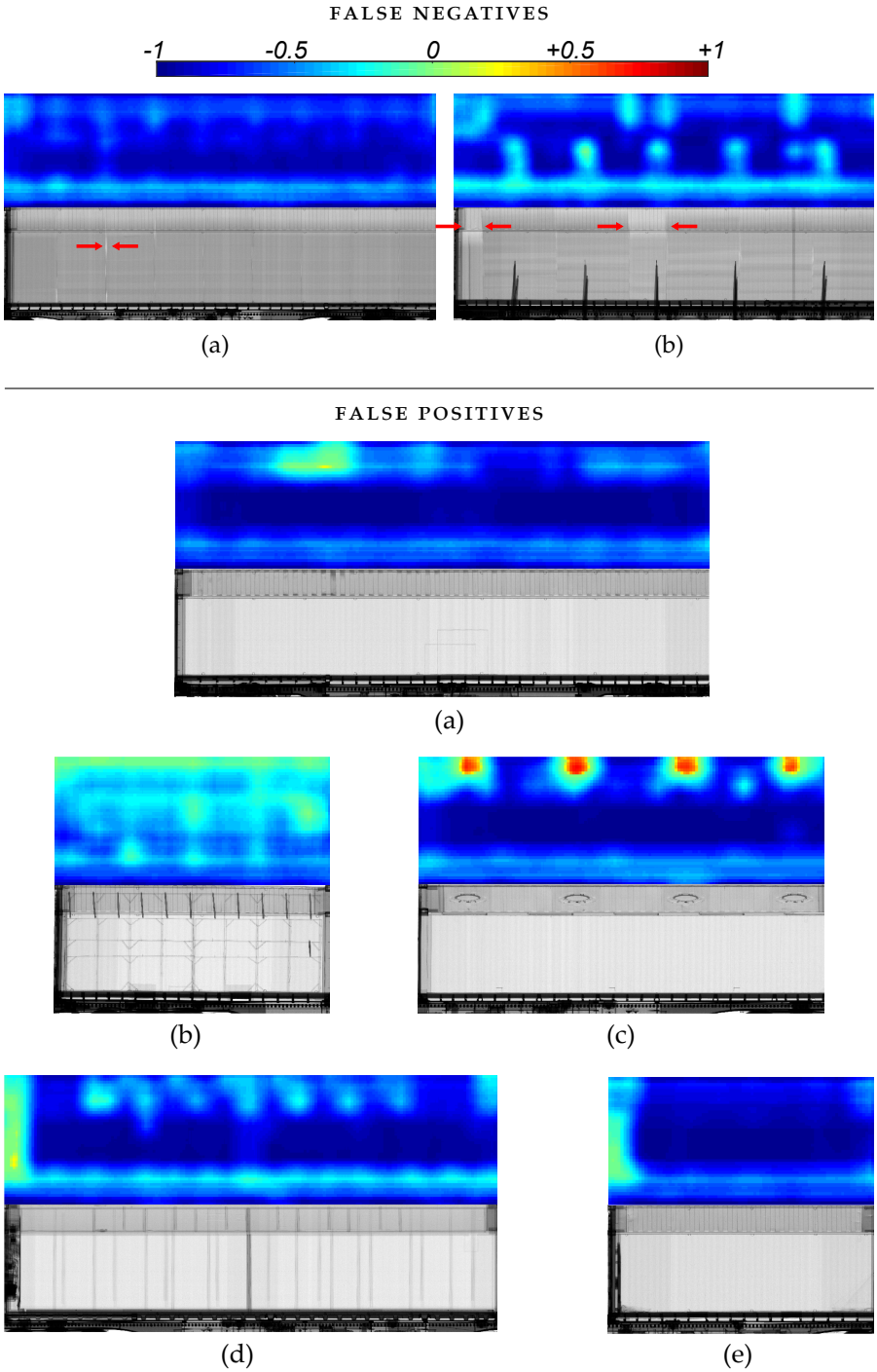


Figure 7.13: Examples of false positives and false negatives corresponding score localisation heatmaps. For the false negatives red arrows indicate cues that the algorithm should pick up to recognised the container as non-empty, this includes slight gaps between cargo in the case of (a) or large spaces in (b). Note that the cargo in these cases occupies the whole container and is very low density, making it difficult to see. The ECV systems has picked up some small signals in these cases, but not enough to trigger any alarms. Not that the thin dense spikes protruding from the floor in (b) are fork-lift pockets and so are correctly classified as empty.

7.4.4 *Testing on difficult TIP examples*

The performance of the system is plotted as a function of the load density and volume in Figure 7.14 (left). Performance is measured as the negative mean log FNR given a 1% FPR. The mean is taken over 10 separate trains of the RF. The blue line indicates the 0.1% FNR contour, and the cyan line indicates a 1% FNR. The density and volume of (i) an average car, (ii) 1 L of water, and (iii) 1 kg of cocaine, have been plotted for reference. As expected the system performance drops as the load gets smaller and less dense, i.e. towards the bottom-left of the plot. The system is able to detect >90% of loads as difficult as 1 L of water and 1 kg of cocaine, with <1% false positives.

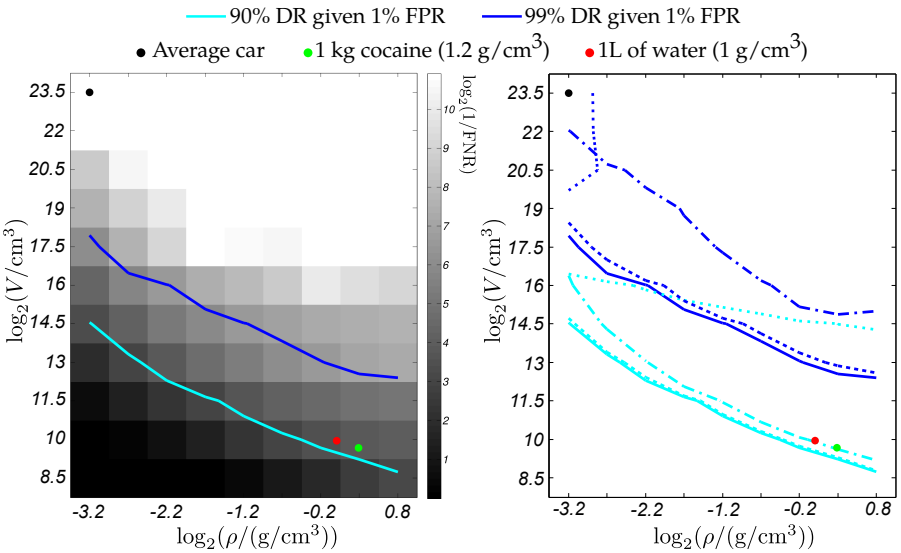


Figure 7.14: *Left*: The system performance (negative log of the FNR given 1% FPR) plotted as a function of the log-volume (log V) and log-density (log ρ) of the load. The blue and cyan lines show the contour for 99% and 90% detection respectively. *Right*: The detection contours when using all features (solid); minus oBIFs (dotted); minus moments (dashed); minus coordinates (dot-dashed). Results show that each feature class contributes to performance.

Figure 7.14 (right) shows contours from three additional classifiers, each excluding one of the classes of features: oBIFs, moments, or window coordinates. From these results it is evident that the classifier that uses all features (as in Section 7.3) performs best, and that each of the three elements of the feature vector contribute to performance. In order, the greatest hit to performance is when oBIFs are dropped, followed by window coordinates, and then intensity moments.

To further understand how the system operates, the Gini feature importance of the different features was computed. Gini feature importance is a

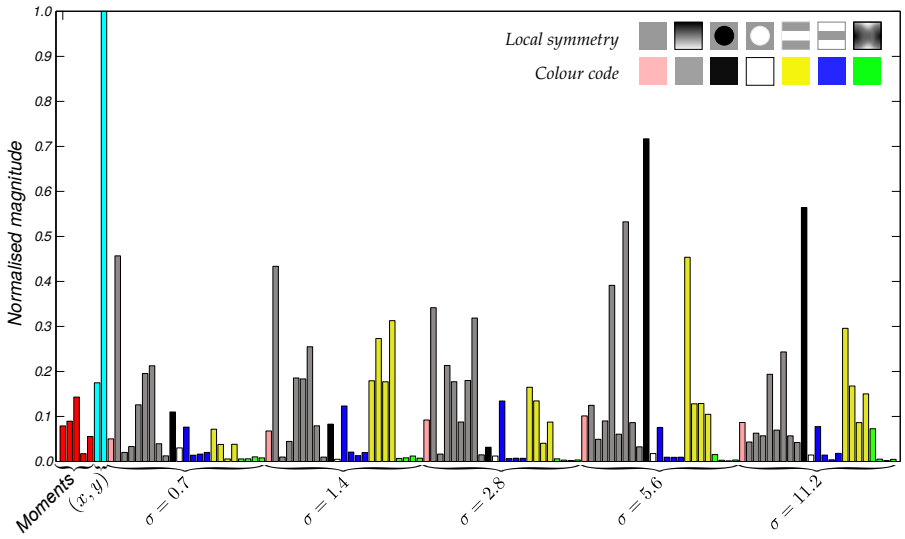


Figure 7.15: The importance (mean decrease in Gini index) of each feature, oBIFs have been colour coded according to their type. The window y-coordinate is most important, it helps the classifier learn the difference in appearance between the floor and roof.

measure of how useful different features are for the RF to make accurate classifications. It relies on computing the Gini impurity for each feature at each node in the RF. The Gini impurity, a computationally efficient approximation to entropy, at a node, measures how well the data is split into separate classes. The Gini impurity for each feature is then aggregated across all nodes in a given tree, and all trees in the Forest yielding an overall measure of feature importance.

The feature importance is shown in Figure 7.15. The feature with largest importance is the window y-coordinate (cyan). This is intuitive because most of the spatial variation in the container occurs in the y-direction. For example, the container changes from roof, to walls, to floor, to underfloor as you go from the top of the image to the bottom. Due to this variation in background appearance the classifier uses the y-coordinate to learn what empty windows should look like at different y-locations in the image. It is also interesting to note the increasing importance of the minima-like oBIFs (black) at coarser scales. This indicates that the system is looking for broad dark blobs in the image. Furthermore, the system finds light line-like (yellow) oBIFs useful, but dark line-like (blue) oBIFs less so. This may be because light lines are usually an indicator of load (e.g. two loads placed next to each other) and the container background has few light lines, whereas both load and container tend to have a lot of dark lines.

Figure 7.16 shows cropped examples of difficult TIP loads (V=1×10^3 cm^3 and ρ=0.9 g/cm^3 i.e. similar to 1 L of water). Note that the system is able to detect difficult TIP loads even where it is difficult to distinguish them from background by eye, such as in Figures 7.16d and 7.16h. The false negatives are difficult to see by eye, especially without the yellow bounding boxes marked. In the false positives the system has incorrectly classified detritus Figure 7.16i, a garment rail (Figure 7.16j), part of a refrigeration unit (Figure 7.16k), and container damage (Figure 7.16l) as non-empty. The example in Figure 7.16i is so rare, that it would be very difficult to train the classifier to ignore it. In Figure 7.16j the window only overlaps with the end of the garment rail and so locally it looks like a small point load. This might be improved by using fuzzy windows (e.g. Gaussian) or a classifier that combines multiple window sizes.

Score localisation heatmaps for true positives are shown in Figure 7.17. It is evident that the container background is given negative heatmap scores, indicating that the classifier successfully recognised empty background patches. In slightly more difficult cases such as the roof bows in Figure 7.17a or the corners of the container, the score is still negative, but the classifier is not 100% confident, there are a small fraction of RF trees that vote that empty patches are non-empty. The TIP-inserted loads all have a positive heatmap value, indicating that the RF is confident that the load is non-empty. For the false negative examples in Figure 7.17e-h, the TIP-concealed loads each lead to a small signal, however it is not large enough to trigger a detection. At least, without reducing the threshold which would result in many more false positives.

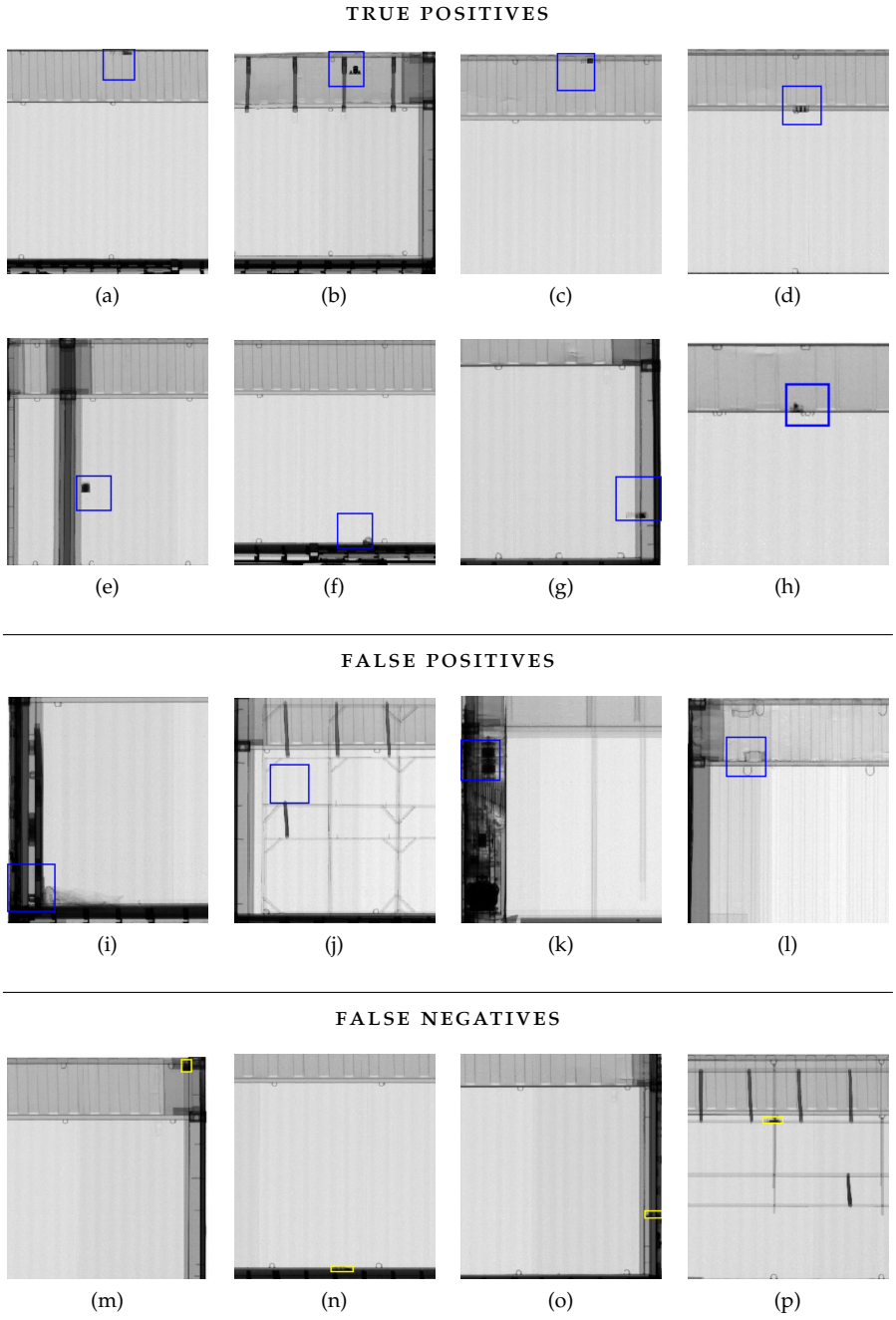


Figure 7.16: Example detections for loads with V=1000 cm^3 and (ρ=0.9 g/cm^3, i.e. similar to 1 L of water. Blue rectangles indicate that a window was classified as non-empty. Images have been cropped. Yellow rectangles indicate false negative load ROIs.

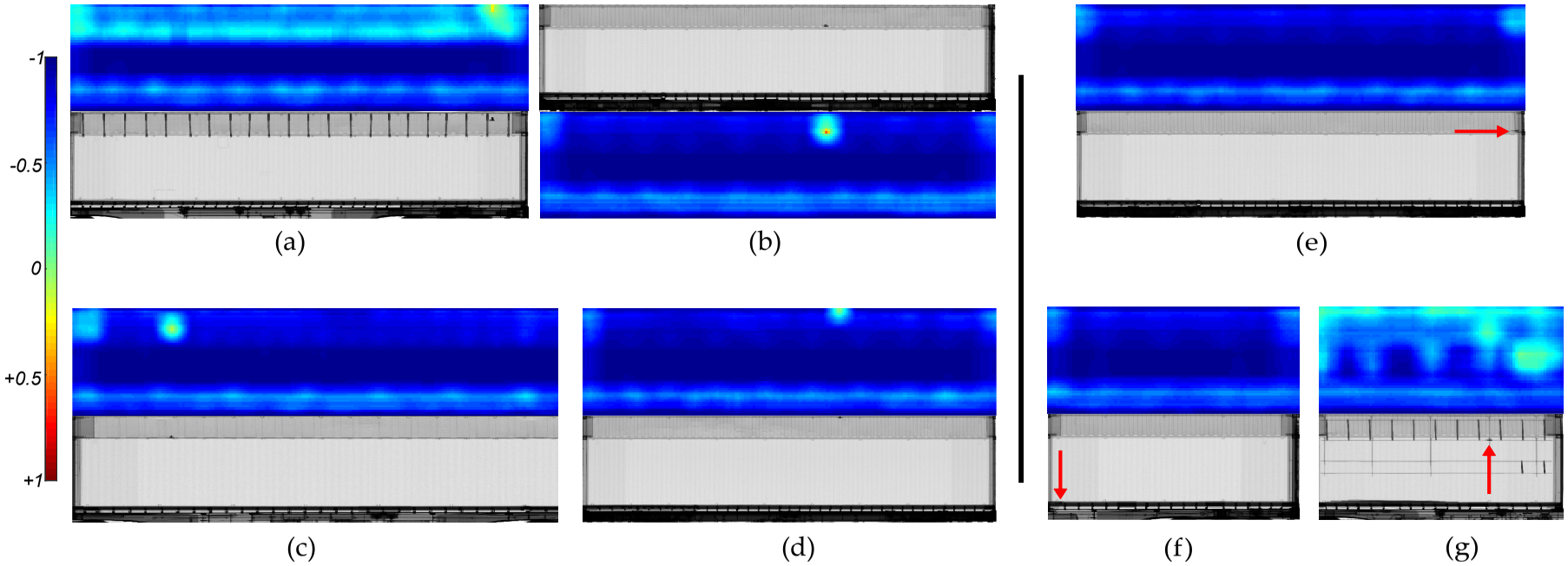


Figure 7.17: Examples of detected Threat Image Projection (TIP) examples (a-d) and false negatives (e-g) and the corresponding localisation heatmaps. In this case the loads are approximately equal in volume and density to 1 L of water.

For each volume-density point on the performance plot in Figure 7.14 (left), an ROC curve can be determined. Figure 7.18 shows ROC curves (True Positive Rate (TPR) versus FPR) for loads similar in volume and density to different masses of cocaine. The density of cocaine has been taken as $1.2\,g/cm^3$. As the mass of cocaine decreases it is evident that the area under the ROC curve, and thus the performance of the system, decreases. The system is able to achieve >93% detection (with a 1% FPR) for loads with mass $\geqslant$1.25 kg of cocaine (and the same density). For smaller masses the performance drops steeply, for example the system is able to detect only ~80% of loads (given 1% FPR) similar to 0.5 kg of cocaine.

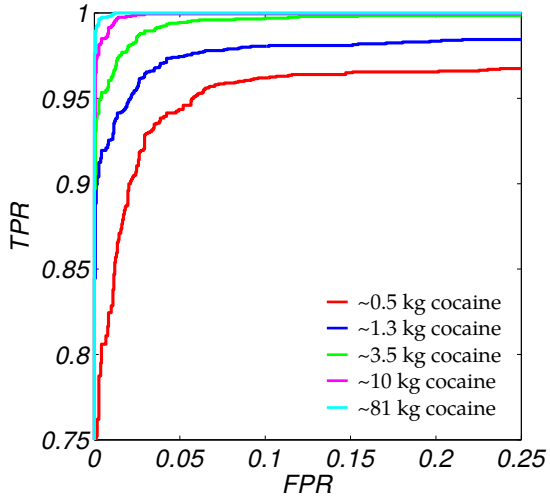


Figure 7.18: Receiver Operating Characteristic (ROC) curves for different masses of cocaine. A cocaine density of $1.2\,g/cm^3$ has been assumed.

Heatmaps can also be employed to study the distribution of the false negatives and false positives across the the container. False positive and false negative heatmaps are shown in Figure 7.19 along with the mean empty container. Note that the x-coordinate is in nearest-container-end coordinates, and that the container height varies between images so that the container floor is usually located between 500<y/Px<600. The false positives are mostly due to refrigeration unit parts that look like load (Figure 7.16k), or port holes. The false negatives mainly occur when the load is projected onto a dense part of the image such as into the container wall, floor or corners. Such examples are difficult to see by eye.

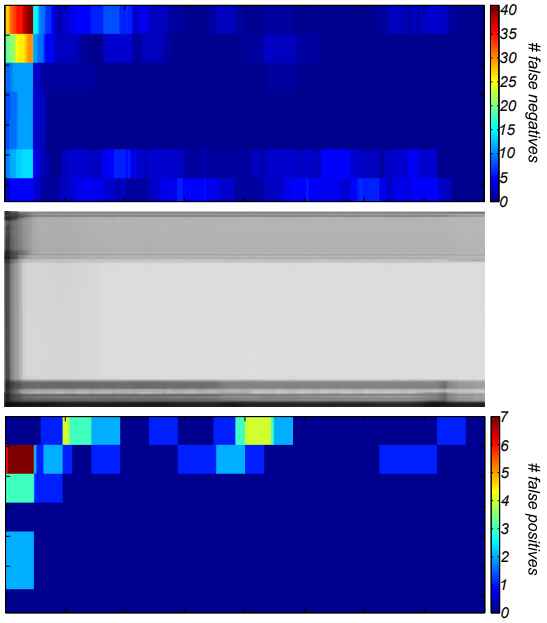


Figure 7.19: *Top*: Heat map of false positive windows. *Centre*: the mean cargo container. *Bottom*: Heat map of false negative windows. The x-coordinate is the window coordinate taken from the nearest container end, and the mean container image was computed accordingly. The y-coordinate is measured from the top of the image.

7.5 DISCUSSION

Declared-as-empty containers are often used to smuggle contraband due to lack of surveillance and ease of access. In historical smuggling cases, amounts of contraband smuggled in declared-as-empty containers has varied from as little as 8 kg to as much as 514 kg [39]. A system for ECV in X-ray images of cargo containers has been proposed and tested. Since there are few examples of difficult loads in the stream-of-commerce, the TIP framework from Chapter 6 was used to generate of realistic difficult non-empty examples. The TIP framework also enables the performance of the system to be assessed as a function of the load volume and density.

The system is based on using a RF classifier to distinguish non-empty and empty windows in the image. For each window, oBIFs [157] and intensity moments are used as feature descriptors. In addition, the window coordinates are used as a feature, allowing the RF to implicitly learn variation in the background (container) appearance as a function of position within the image.

The system is able to achieve performance of 99.3% detection and 1% false positives on real stream-of-commerce images, and thus outperforms the work of Orphan et al. [66]. Moreover, it achieves >93% detection whilst

raising 1% false alarms on difficult TIP examples that are similar in volume and density to 1.25 kg of cocaine. The majority of false negatives are when the load is placed in a dense region of the container, such as in the floor. Such examples are difficult for even humans to see by eye, and future work should focus on improving performance in such cases.

8

DUAL-ENERGY AUTOMATED THREAT DETECTION

In this chapter a method for Automated Threat Detection (ATD) is proposed. The method exploits dual-energy measurements to implicitly learn about material characteristics to suppress false alarms. Several methods for exploiting dual-energy were explored, including three based on material discrimination methods in the literature, and three novel variants. Thus it serves as a comparison between different material discrimination methods. The best performing system makes a 100-fold improvement on the false alarm rate over the prior work [10]. The system is trained and tested on Threat Image Projection (TIP), however the Empty Image Projection (EIP) validation proposed in Chapter 6 is employed to test whether the system can exploit TIP artefacts to boost performance.

In this chapter, the term 'small metallic threats' is used as the research results should not be easily discoverable by keyword searching. However, the smallest of the threats in question are similar in form to hand drills, whilst the largest are similar in length to a garden spade.

8.1 MOTIVATION

In recent years the threat from Mumbai-style terrorist attacks using Small Metallic Threats (SMTs) on soft targets has increased. This is evidenced by the attacks in Paris and on the beaches of Tunisia. In such attacks, a few perpetrators, by acting as a well-organised unit striking simultaneously against unprotected civilian targets in urban areas, can inflict mass fatalities and casualties. To prevent such attacks, it is necessary to make it more difficult for would-be terrorists to obtain SMTs. In countries, where SMTs are well controlled or illegal to possess if too powerful, would-be terrorists rely on creating sophisticated smuggling networks, which is expensive, or on obtaining SMTs from the existing pool.

They could be obtained by theft, loaned from a criminal associate, or purchasing from the black market.

This chapter builds on the work by Jaccard et al. [10], where it was found that containers containing a single SMT could be detected in 90% of cases whilst raising 6% false positives. Additionally, Jaccard et al. [10] found that an oBIF-based system similar to that used for Empty Container Verification (ECV) in Chapter 7 did not perform well. This work used a single energy X-ray image acquired with a 6 MeV energy cut-off. In this chapter, it is investigated whether the addition of an image acquired at an energy cut-off of

4 MeV can be exploited to improve detection performance. In theory, the use of two energies should provide a Convolutional Neural Network (CNN) with some information about material types that can help to reduce false alarms. Whether this potential information can be unlocked in CNNs is unclear.

To best knowledge, no other research has been performed on how to utilise *raw* dual-energy measurements in a CNN architecture, in any field. The closest example was the use of colour-coded baggage images in aviation security by Açkay et al. [127], where the false colour information essentially provides a denoised pseudonym for dual-energy information across three colour channels.

8.2 CONVOLUTIONAL NEURAL NETWORKS

The historical roots to CNNs can be traced back to the simple model of the neuron by McCulloch and Pitts [171]. Their artificial neuron is a simple mathematical model of the biological neuron, and implements many of its important features. The basic principles of an artificial neuron are that:

(i) it has a number of inputs each with its own weighting;

(ii) the weighted inputs are summed and compared to a threshold (bias);

(iii) if the threshold is exceeded then the neuron fires.

By building networks of neuron layers, each layer feeding into the inputs of the next, these basic building blocks could be used to perform simple pattern recognition tasks. Early researchers could hand-craft the weights and biases to perform the desired task.

However, researchers found that such simple nets could not perform more sophisticated tasks. So the perceptron and Multi-Layer Perceptron (MLP) were eventually invented. In the MLP, rather than hand-crafting the weights and biases, they were learnt directly from training data. The training required the use of backpropagation to determine at each training step, how much the weights and biases in the network should be updated to reduce the overall error of the network. The backpropagation requires a slight modification to the activation function; rather than a simple threshold a continuous differentiable function is required such as a sigmoid function. The MLP had some success in image recognition tasks, however it was clear that the network suffered from the 'curse of dimensionality', due to the full con-

nectivity between neurons. Since each network has a very large number of parameters they were difficult to train and prone to overfitting.

CNNs have recently become the state-of-the-art for a range of visual tasks. They are a type of Deep Learning method, and differ from traditional computer vision methods since they learn an end-to-end mapping from the raw image values to the image classification. As part of this, they learn image features, instead of requiring them to be hand-crafted by researchers, as well as their optimum combination to make a classification.

CNNs are variants of MLPs and are also biologically inspired. They differ in two important ways: (i) local connectivity and (ii) shared weights. Local connectivity implements the idea of receptive fields, which is motivated by biological vision [172]. They exploit spatially local correlation by enforcing local connectivity between neurons of consecutive layers. Each neuron analyses a small region (receptive field) of the preceding layer, and the receptive fields are arranged in a lattice; this is equivalent to a convolution. Only one set of weights and a bias has to be learnt and is shared between all of the neurons in the lattice (i.e. a single filter). The convolution operation means that there are fewer parameters to learn than in a fully connected network; the weights are shared.

A CNN consists of multiple convolutional layers and often ends with multiple fully-connected layers and a classification layer. The first layer has been shown to learn simple filters, similar to Gabor filters. The next layer will be some combination of these simple filters. The level of complexity in the features increases as more convolutional layers are added. By the later layers, it has been visualised that, complicated features for whole objects, such as faces or vehicles, begin to emerge.

Another important feature of CNNs are the use of Rectified Linear Units (ReLU) and pooling layers (e.g. *max pooling*). ReLUs are the most commonly used activation functions in CNNs and are defined as $f(x) = \max(0, x)$. In *max pooling* layers, the preceding layer is split into non-overlapping regions and the *max pooling* layer outputs the maximum in each region. The motivation is that the operation preserves the most important features, but progressively reduces the number of parameters in the network and the amount of computation.

CNNs can be prone to overfitting particularly if trained on insufficient data. Thus, a number of regularisation methods have been developed, with Dropout [173] being the most popular until recently. In Dropout, a different

random selection of nodes from fully-connected layers are silenced in each training pass. Recently, Ioffe and Szegedy [174] have introduced *batch normalisation*, which fixes the mean and variance of input distributions at each layer. It acts as a form of network regularisation and can allow faster network training and higher model accuracy.

LeCun et al. [175] give a complete overview of the fundamental concepts of CNNs.

8.3 EXPLOITING DUAL-ENERGY

Material discrimination was introduced in Section 4.2.4. Here we move away from the terminology in the material discrimination literature, and refer to the mappings of the H and L images as 'feature spaces' to match common terminology in computer vision.

Cargo material discrimination works by performing image measurements at two different energies. Based on these two images, simple features can be computed such that different material atomic numbers (Z) are partitioned in feature space. These methods operate on the radioscopic transparency, T. Given the equation of image formation

$$I = \int I_0(E) e^{-\mu(E,Z)\tau} dE, \tag{8.1}$$

the transparency is defined as

$$T = \frac{\int I_0(E) e^{-\mu(E,Z)\tau} dE}{\int I_0(E) dE}. \tag{8.2}$$

In these two equations E is the photon energy, $I_0(E)$ is the initial intensity of photons emitted from the source, $\mu(E, Z)$ is the attenuation co-efficient for a material with atomic number Z for incident photons with energy E, and τ is the material thickness. To keep notation simple for the rest of this chapter, the high energy and low energy transparencies are denoted by H and L, respectively.

In their seminal work for material discrimination in cargo, Ogorodnikov and Petrunin [5] use the log-ratio R and $1/H$ as features. The log-ratio is defined as

$$R = \frac{\log(H)}{\log(L)}. \tag{8.3}$$

This is motivated, in part, by the observation that for a monochromatic beam of energy E_γ such that

$$I_0(E) \propto \delta(E - E_\gamma), \tag{8.4}$$

where $\delta(\_)$ is the Dirac delta function, then the ratio of logs becomes

$$R \propto \frac{\log(e^{-\mu(E_H,Z)\tau})}{\log(e^{-\mu(E_L,Z)\tau})} = \frac{\mu(E_H,Z)}{\mu(E_L,Z)}. \tag{8.5}$$

So for a given material with atomic number Z, R is unique to that material and does not depend on the thickness τ. R can therefore be used to discriminate materials. However, for cargo screening, the X-ray photons are generated by the Bremsstrahlung process, and thus the high and low energy beams are not monochromatic, but a continuous spectra up to some cut-off energy equivalent to the energy of the electrons used to generate the X-rays. In this polychromatic regime, material discrimination is still possible, but the log-ratio R is not unique for a particular material.

Analytic material curves in the Ogorodnikov and Petrunin [5] feature space are given in Figures 8.1c and 8.1f, for monochromatic and polychromatic cases, respectively. No noise has been modelled in these examples. In the monochromatic case, the curves are well separated in feature space, however, in the polychromatic case, they overlap for small $1/H$ values (as materials become thin). The material curves are also bunched closer together, meaning that noise in the imaging system can more easily lead to material misclassification. Ogorodnikov and Petrunin [5] note that material discrimination is also difficult in the case of thick materials since noise begins to dominate the image.

There are two other methods of computing features in the literature. These are known as the α-curve [97, 98] and H−L curve [142] methods. The α-curve method computes the features

$$\alpha_1 = -\log(H) \tag{8.6}$$

$$\alpha_2 = -\log(H) + \log(L), \tag{8.7}$$

and the H−L curve method simply uses H and L as the features. Analytic material curves for these approaches are also given in Figure 8.1. Note that in the polychromatic case the H−L curves are bunched very close together compared to the α-curve and R-curve methods, and so one would expect this method not to perform as well.

To form an RGB image, authors typically perform a system calibration by scanning known materials of varying thickness. The calibration image can be use to determine how feature space should be partitioned such that pixels can be classified into material Z groups. Each group is assigned a different

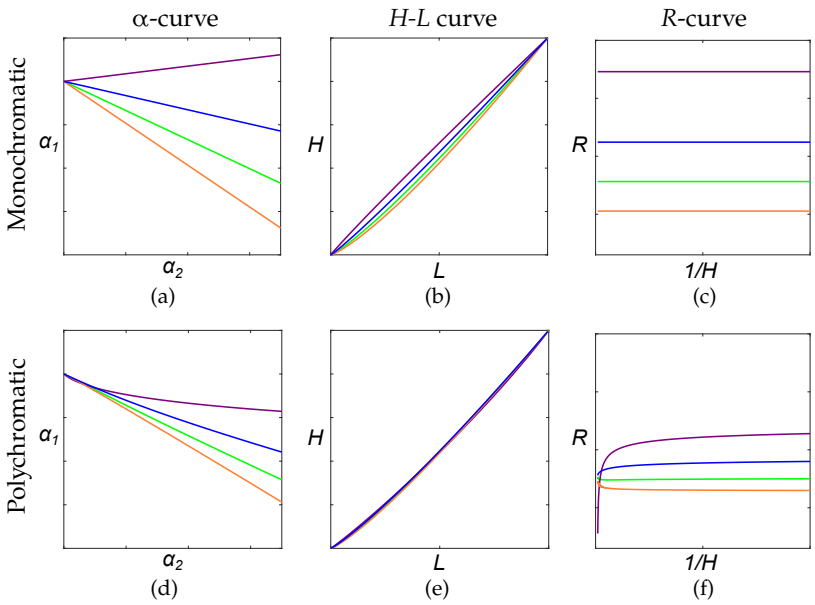


Figure 8.1: Material curves in feature space for three different material discrimination methods in the literature. The materials, include: Boron (Z=5, orange); Aluminium (Z=13, green); Iron (Z=26, blue); and lead (Z=82, purple). Curves were computed analytically using Equation (8.1), assuming a noiseless imaging system, and using attenuation coefficients from the NIST XCOM database [62].

hue in the coloured X-ray image. Ogorodnikov and Petrunin [5], found that system noise leads to a noisy RGB image and spatial information is required to improve the quality of the RGB image. The authors apply a simple segmentation algorithm and each segment is labelled as the average material over the pixels in that segment. Since the original work of Ogorodnikov and Petrunin [5], which used a controlled laboratory set-up, other researchers have failed to replicate their accurate results for commercial systems. Some authors have focussed on detecting only high-Z materials, which can be indicative of nuclear material or adversarial shielding, citing that multi-class material discrimination is infeasible due to the levels of noise in commercial systems [74].

It is possible that machine learning approaches, and in particular deep CNNs, can learn to perform material discrimination based on dual-energy measurements and spatial information in the image. However, a major hurdle to achieving this is the difficulty obtaining large datasets of different materials with pixel-level labelling. The aim of this work is to implicitly learn material discrimination in order to boost performance in ATD.

For this work, the three feature spaces from the literature (Section 4.2.4) and three novel variants, are explored.

The first variant is referred to as the $\Sigma{-}\Delta$ curve method. This is similar to the H−L curve method [142], but rather than using H and L as features, Σ=H+L and Δ=H−L are employed. This approach yields a larger separation between material curves in the feature space, and so one would expect better material discrimination as a result. The material curves for the $\Sigma{-}\Delta$ curve method are given in Figures 8.2b and 8.2e.

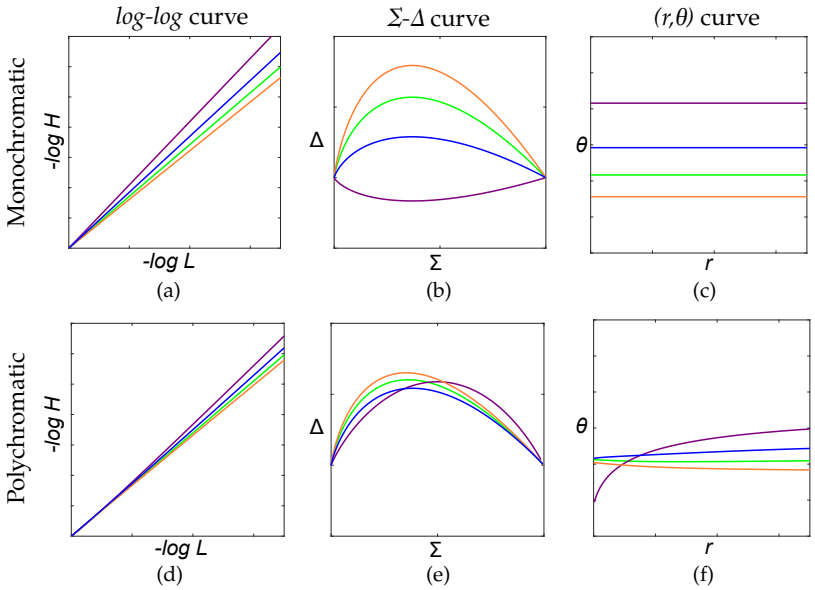


Figure 8.2: Material curves in feature space for three novel variants of material discrimination methods in the literature. The materials, include: Boron (Z=5, orange); Aluminium (Z=13, green); Iron (Z=26, blue); and lead (Z=82, purple). Curves were computed analytically using Equation (8.1), assuming a noiseless imaging system, and using attenuation coefficients from the NIST XCOM database [62].

The second variant is the *log-log* method, which is similar to the α-curve method, but uses just $-\log H$ and $-\log L$ as features. This is convenient since, when combined with the H−L curve method, it is a straightforward dual-energy generalisation, i.e.

$$\{-\log H, H\} \rightarrow \{-\log H, -\log L, H, L\}, \tag{8.8}$$

of the single-energy system [10], which uses H and $-\log H$ as CNN input channels. This provides a good baseline for dual-energy networks.

The third variant is the (r, θ) method. This is derived by transforming the *log-log* method into polar co-ordinates:

$$r = \sqrt{(\log L)^2 + (\log H)^2} \tag{8.9}$$

$$\theta = \operatorname{arctan2}(\log H, \log H). \tag{8.10}$$

In a monochromatic, noiseless imaging system, it would be possible to discriminate materials using just θ. $\operatorname{arctan2}(\_,\_)$ *is the four-quadrant inverse tangent.*

In the monochromatic case, this gives θ-values that are constant as a function of r. In the polychromatic case, the curves appear similar to the R-curve method, but there is a better separation between materials at small r-values.

8.3.1 *Dual-energy network architectures*

The CNN architectures are based on the 19-layer very deep networks first described by Simonyan and Zisserman [132], and have been shown to work well in single-energy ATD [10]. The main modification to the original Simonyan and Zisserman [132] architecture, is the addition of *batch normalisation* layers [174]. In total, the networks contain 16 convolutional layers, 3 fully-connected layers, and a *softmax* layer to obtain the confidence that a patch contains an SMT. The input channels are of dimension 256×256.

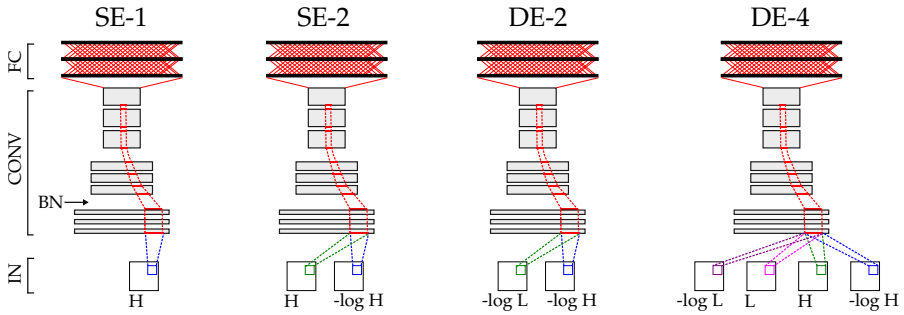


Figure 8.3: Illustration of the single-energy one-channel (SE-1) and two-channel (SE-2) CNN architectures used in the prior work [10] and operating on single-energy inputs (IN), and the two-channel (DE-2) and four-channel (DE-4) architectures employed in this work. Each network has a number of convolutional layers (CONV) interspersed by batch normalisation (BN) layers. The final convolutional layer feeds into the fully-connected (FC) layers. The networks used in this work consist of 16 convolutional layers and 3 fully-connected layers (not depicted in this figure).

Networks were investigated that have (i) two dual-energy input channels, and (ii) four dual-energy input channels. It was found in previous work that using separate input streams rather than channels, does not improve performance [10]. Illustrations of these network architectures are given in Figure 8.3. At first, each of the material discrimination methods are assessed as two-channel network. For each method, the H and L images are transformed into the feature space of that method, and each feature image fed into the CNN as a separate input channel. Next the best performing two-channel inputs are combined to construct a four-channel network, to see if performance can be further improved. All inputs are rescaled to the range $0-255$ and cast to 8-bit.

It is possible that a slight performance improvement is possible using 16-bit, but network training takes longer.

In all experiments, the CNN hyper-parameters weight decay and momentum are fixed at 10^{-4} and 0.9, respectively. In most experiments, the learning rate was decreased from 10^{-3} to 10^{-6} over the course of 30 epochs, this was reduced (and number of epochs increased) in experiments, where training was erratic. The mean image computed across the training set was subtracted from each input image. The training and evaluation were performed on an NVIDIA TITAN X, which could handle a batch size of 50 images when using four input channels. All CNNs were implemented and trained using the `MatConvNet` toolbox for `MATLAB` [176].

8.3.2 *Data, pre-processing and augmentation*

The same Stream-of-Commerce (SoC) dataset, captured by a Rapiscan Eagle® R60, as in Chapter 7 was employed.

Of the 1.2×10^5 benign images, the following dataset splits were employed:

- 1×10^4 full SoC images were reserved for testing. Of these 5×10^3 were kept for the benign class, and a single SMT was projected into the other 5×10^3 as the threat class;
- From the remaining 1.1×10^5 images, 6.4×10^5 256×256 patches were sampled for training. A small subsample of 1.25×10^4 patches is reserved for computing the validation error when training the CNN, and this is kept disjoint from the training patches.

The SMT staged threat examples were collected by Rapiscan Systems.

The SMT images were acquired separately. In total, 700 SMT images were collected for a variety of different poses, types and models. The SMT instances were extracted from the original scans to form a dual-energy TIP library. To generate *de-novo* examples for training an SMT was randomly selected from the TIP library and projected into the background patch. Projecting the same SMT instance into different images results in vastly different appearances due to the translucency property of X-ray images and was shown to be indistinguishable from real threat imagery in Chapter 6. The dataset was made more diverse by the injection of realistic variations including translations and flipping. SMT instances were kept disjoint between the training, validation, and testing datasets.

Images were pre-processed according to Chapter 7.

8.3.3 *Training schemes*

In this work, two modifications to the CNN training routine are tested. First, TIP and data augmentation are performed on-the-fly, meaning that even if background patches are reused, each instantiation will use TIP with a different SMT projected under different random conditions. This means that there is a lot more variation in the threat class used in training. In the prior work [10], training patches were computed prior to training and saved to disk.

Second, a 'spot-the-difference' training scheme is proposed. In each training batch, each synthesised example in the threat class is identical to one in the benign class except for the projected SMT. The rationale for this is that it can improve CNN training since the network can quickly learn to focus efforts on learning features for SMTs since the only discriminator between the two classes is the presence of the, often heavily shielded, SMT. However, this approach could also be disadvantageous since the network is exposed to less variation.

8.3.4 *Performance evaluation*

The performance of the different systems was evaluated on full-sized container images with a single SMT projected into the container or cargo. A sliding window approach is adopted to analyse the whole image. Windows of size 256×256 pixels are sampled with a stride of 64 pixels in both the vertical and horizontal direction. For each window a confidence is computed according to the output from the *softmax* layer. A confidence score for the whole image is computed by taking the maximum of the window confidences.

Performance is assessed in terms of the Area Under the Curve (AUC), and false positive rates for fixed detection rates of 90% (FPR90), 95% (FPR95) and 99% (FPR99).

To understand the full-image classification results, a heatmap of window confidences is computed. This allows one to locate false positive or detection signals, and is potentially a useful visualisation for operators to quickly identify threats. However, since the heatmaps are computed by sliding a 256×256 window across the image, the resolution of the heatmap is poor. It

would be beneficial to have a method of localising the SMT signal within this region.

To this end, the method used by Zeiler and Fergus [177] was implemented, for determining the strongest cues in a window that the CNN has used to detect SMTs. This works, by sliding a small occluding window across the 256×256 region. For each position of the occluder, the CNN score is computed. If the occluder is blocking a part of the image that provides very strong cues to the CNN, then this results in a much lower score. Thus a heatmap can be constructed which has low scores corresponding to the most SMT-like image parts, and in this way the detected SMT can be localised. This is particularly useful for the smallest SMTs when hidden amongst complicated background structure, or localising false positive detections.

8.4 RESULTS

First the new training schemes are tested on a single-energy network and compared to Jaccard et al. [10], then the best scheme is adopted for training and testing dual-energy networks.

8.4.1 *Training schemes*

Two-channel single-energy (SE-2) networks with inputs $\{-\log \mathrm{H}, \mathrm{H}\}$, were trained with (i) on-the-fly training without using spot-the-difference examples, and (ii) on-the-fly training whilst using spot-the-difference examples. The results are summarised in Table 8.1. The 95% confidence intervals of each of the performance metrics were determined by taking a bootstrap sampling of the image scores and using the bias corrected and accelerated percentile method. Analysis was performed with the same epoch for each network to allow fair comparison, and the epoch was chosen such that the validation and training errors had plateaued for both networks.

OtF?	**StD?**	**AUC**/%	**FPR90**/%	**FPR95**/%	**FPR99**/%
N	N	97.0	6.00	–	–
Y	N	98.5 [98.3, 98.7]	1.66 [1.34, 2.02]	8.83 [8.02, 9.65]	30.9 [29.6, 32.1]
Y	Y	98.5 [98.3, 98.7]	1.92 [1.55, 2.31]	7.73 [7.07, 8.53]	34.4 [33.1, 35.8]

Table 8.1: Results on full cargo images, for the different training approaches: StD and OtF. The SE-2 architecture with inputs $\{-\log \mathrm{H}, \mathrm{H}\}$ was used, as in Jaccard et al. [10]. The values in square braces correspond to the 95% confidence interval, obtained by bootstrap sampling with 1000 samples.

In both tests, the performance was significantly improved over the original off-the-fly training scheme. However, using spot-the-difference did not provide a statistically significant improvement over using just on-the-fly. Although, intuitively, the spot-the-difference method gives help to the CNN by allowing it to focus weight updates on learning features for SMT detection, it does not yield a boost in performance. This is possibly due to the reduced variability of images that the network is exposed to. A benefit of using the spot-the-difference scheme is that batch processing time is improved. This is because only half the number of patches have to be read from disk since the same patches are used for the benign class and generation of the TIP examples.

Overall, on-the-fly training improved the AUC by +1.5%, and the false positive rate given 90% detection improved from 6% to 1.66%, and is thus a significant improvement over the prior work [10].

8.4.2 *Two-channel dual-energy networks*

Next it is investigated how each material discrimination method performs when used alone in a two-channel network. Table 8.2 shows the results, complete with 95% confidence intervals as before. The H−L curve method, as expected from observing the material curves in Figure 8.1, does not perform well in comparison to the other material discrimination methods. However, there is +13% improvement on the single-energy network operating on just on H, and so the H−L curve method of material discrimination does provide some benefit. Similarly, the *log-log* method, yields a +2.8% improvement in AUC and 9-fold improvement in FPR90 over its single-energy analogue (operating on $-\log H$).

The Σ−Δ novel variant offers a significant improvement on the H−L curve method, boosting the AUC by >4% and giving a 4.5-fold improvement on FPR90. This is expected from comparing the material curves, in Figures 8.1e and 8.2e, since the latter has better separated materials in feature space. However, it is surprising that the CNN cannot learn simple linear combinations of H and L inputs to match the performance of Σ−Δ.

The best performing material discrimination method was the $\{r, \theta\}$ method. It outperformed all of the other DE-2 methods, yielding an AUC of 99.1% and FFPR95 of 0.62%. It was found that the R-curve method failed to converge, so results are unavailable. This is likely a result of the ratio of logarithms,

Arch.	Inputs/method	AUC/%	FPR90/%	FPR95/%	FPR99/%
SE-1	$\{H\}^\dagger$	89.0	47.0	–	–
	$\{-\log H\}^\dagger$	96.0	9.00	–	–
SE-2	$\{-\log H, H\}^\dagger$	97.0	6.00	–	–
	$\{-\log H, H\}$	98.5 [98.3,98.6]	1.66 [1.34,2.02]	8.83 [8.02,9.65]	30.9 [29.6, 32.1]
DE-2	$\{H, L\}$	92.2 [91.7,92.7]	34.4 [33.2,35.8]	52.2 [50.8,53.6]	68.5 [67.3,69.8]
	$\{\Sigma, \Delta\}$	96.8 [96.5,97.1]	7.70 [7.00,8.45]	24.7 [23.5,25.9]	55.9 [54.5,57.4]
	$\{-\log H, -\log L\}$	98.8 [98.6,99.0]	1.04 [0.78,1.34]	3.39 [2.92,3.92]	**17.8** [16.8,18.9]
	$\{\alpha_1, \alpha_2\}$	98.6 [98.4,98.8]	1.72 [1.39,2.09]	7.81 [7.09,8.63]	28.3 [27.0,29.5]
	$\{r, \theta\}$	**99.1** [98.9,99.3]	**0.62** [0.42,0.88]	**2.18** [1.83,2.63]	**17.8** [16.8,18.9]
DE-4	$\{-\log H, -\log L, H, L\}$	99.2 [99.1,99.4]	0.57 [0.37,0.80]	2.38 [1.97,2.86]	19.4 [18.4,20.7]
	$\{-\log H, -\log L, \Delta, \Sigma\}$	**99.5** [99.4,99.6]	0.08 [0.02,0.18]	**0.36** [0.22,0.57]	**15.5** [14.5,16.9]
	$\{\alpha_1, \alpha_2, \Delta, \Sigma\}$	**99.5** [99.3,99.6]	0.06 [0.02,0.16]	0.86 [0.63,1.18]	18.4 [17.4,19.5]
	$\{r, \theta, \Delta, \Sigma\}$	**99.5** [99.3,99.6]	**0.02** [0.00,0.12]	0.38 [0.24,0.58]	18.3 [16.8,18.9]

† results from Jaccard et al. [10]

Table 8.2: Quantitative results for all experiments. SE-N indicates single-energy architecture with N input channels, DE-N indicates dual-energy architecture with N input channels. The values in square braces correspond to the 95% confidence interval, obtained by bootstrap sampling with 1000 samples.

and the difficulty of finding a sensible range to constrain the image pixels to. The $\{r, \theta\}$ method deals with this by using the $\arctan2(\_, \_)$ function.

Overall, using dual-energy yields a $+0.6\%$ improvement in AUC and improves FPR90 from 1.66% to 0.62%, for the same network configuration (DE-2 versus SE-2). These improvements are statistically significant according to the confidence intervals in Table 8.2.

8.4.3 *Four-channel dual-energy networks*

The baseline four-channel dual-energy method, with inputs $\{-\log H, -\log L, H, L\}$, performs significantly better than its single-energy analogue with inputs $\{-\log H, H\}$. The AUC is improved from 98.5% to 99.2% and there is approximately a 3.5-fold improvement in FPR90. However, the method does not perform significantly better than the two-channel $\{r, \theta\}$ method. When swapping the $\{H, L\}$ channels with $\{\Sigma, \Delta\}$ channels, there is a large reduction in false positives, and the network now outperforms all DE-2 methods.

Of the four-channel networks tested, the $\{-\log H, \Delta, -\log L, \Sigma\}$ inputs performed best across most performance metrics. However, there is no significant difference between the AUCs of the different methods. It seems difficult to push beyond an AUC of 99.5% using dual-energy inputs.

In total the improvement in AUC using dual-energy as opposed to single-energy (both using on-the-fly) is $+1.0\%$, and with the new on-the-fly training scheme performance the improvement $+2.5\%$ over Jaccard et al. [10]. The false positive rate given 90% has improved from 6% to 0.08% by use

of dual-energy and the on-the-fly training. Another way of expressing this difference is that one can turn the detection rate up by +5%, but at the same time improve false alarms by over an order of magnitude.

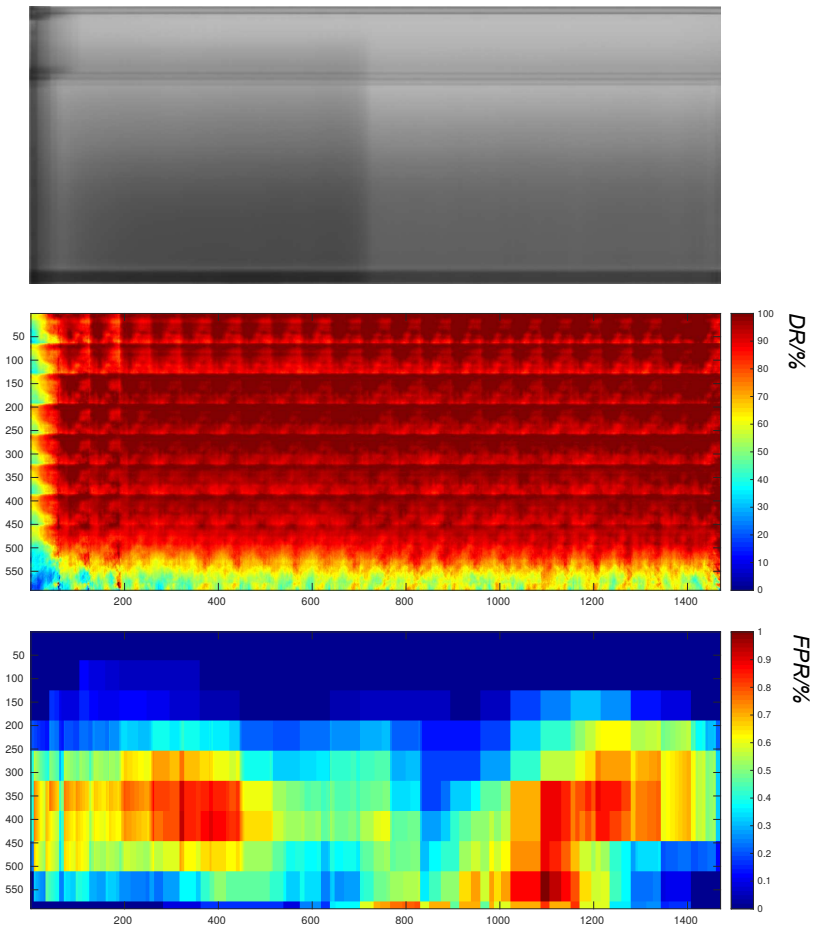


Figure 8.4: Heatmaps of detection rate and false positive rate as a function of container location. The greyscale image is the mean image across the test set.

In Figure 8.4 heatmaps are shown of the false positive rate and the detection rate as a function of location within the container. The detection rate heatmap was constructed using the *threat mask* for each TIP SMT. For each pixel in the heatmap the detection rate was computed based on the threats masks that overlapped with that pixel. For each SoC image, a SMT was projected into each window to build sufficient statistics. The heatmap of the false positive rate was constructed using all image windows in the benign test set. It is evident that the detection rate is close to 100% throughout the cargo region of the container, however it drops significantly when SMTs are projected into the bottom corner or close to the walls of the container. This could be because there are fewer edge and corner patches in the training set, and because these regions are dense, making the SMT signature less visible.

The highest false positives rates appear in the cargo region of the image, indicating that the majority of false positives are triggered from cargo and not from the container structures.

8.4.4 *Classification examples*

In this section, some example classification results are provided for the four-channel network which combines the *log-log* and $\Sigma-\Delta$ methods. The detection threshold was tuned to give a 95% detection rate. Figure 8.5 gives example detections. Some of these examples are very difficult for even humans to detect in a zoomed-view, so localisation heatmaps [177] are used to indicate where the CNN is picking up strong SMT cues.

The strongest cues tend to be on the part of the SMT which is most visible in the image. The fact that the cues are well-localised to the SMT, and not diffuse over the projected patch, provides evidence that the network is not picking up strong cues from potential TIP artefacts that are not located on the SMT itself. This is particularly noticeable in Figures 8.5b and 8.5c, where the majority of the SMT is shielded by very dense cargo, and the strongest cues are located on the small parts of the SMT that are unshielded. In Figures 8.5a and 8.5e, the SMTs are densely shielded, however they are still visible in the log-transformed image. In Figure 8.5d, the SMT has been concealed on complicated background cargo, which makes it difficult to locate by eye.

Figure 8.6 shows examples of false positives. In Figure 8.6b the strongest cues are located at the junction between the floor and two parallel metal tubes. In this example, the false positive appears locally as SMT-like and is also made out of a similar material. In addition, the detected objects in Figures Figure 8.6a, Figure 8.6c and Figure 8.6d, appear similar to components of SMTs, and so the false positives are understandable in these cases.

In Figure 8.7, examples of false negatives together with the CNN inputs for the SMT patch are given. In both cases the SMT cannot be seen, by eye, in the inputs and thus there is very little information for the network to work on. Detection in these cases are very difficult even for humans.

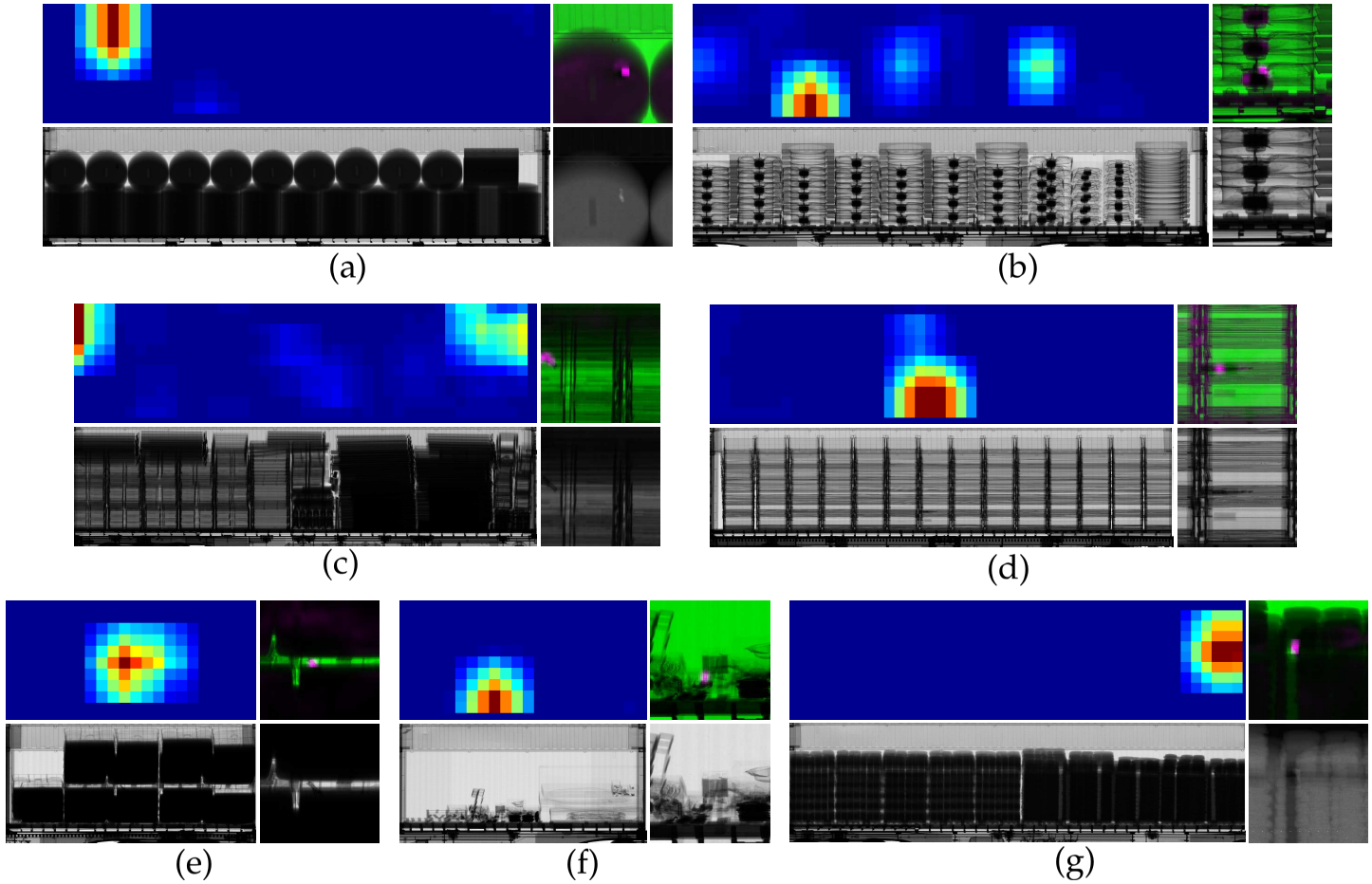


Figure 8.5: Examples of SMT detections in full cargo containers. For each example, going clockwise from top left: (i) score heatmap - red indicates high confidence of SMT; (ii) localised heatmap overlaid on raw image patch - pink indicates strong SMT cues picked up by CNN; (iii) raw or log-transformed image patch depending on which is easiest to see SMT; (iv) original raw image.

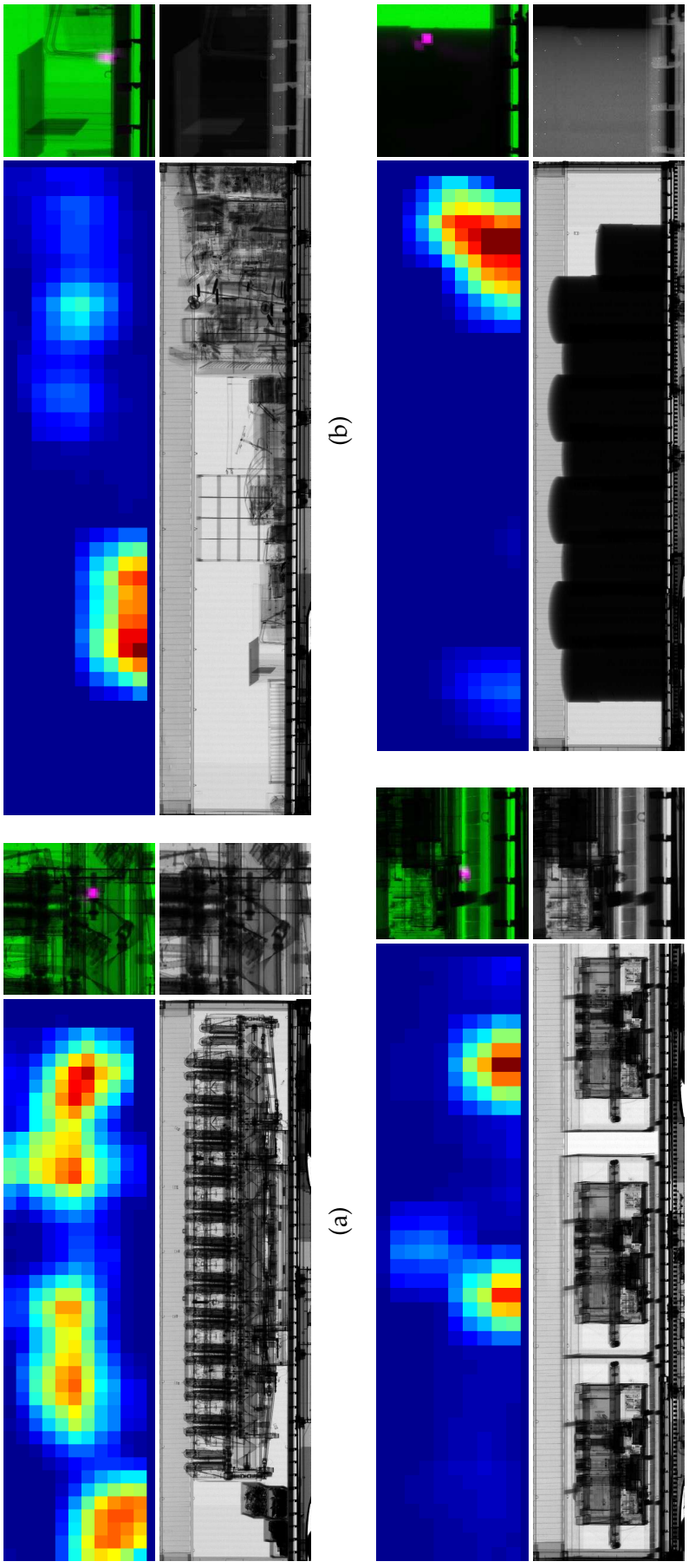


Figure 8.6: Example false positives. For each example (a-b), going clockwise from top left, are displayed: (i) score heatmap - red indicates high confidence of SMT; (ii) localised heatmap overlaid on raw image patch - pink indicates strongest SMT cues picked up by CNN; (iii) log-transformed image patch; (iv) original raw image. In each case the localised heatmap shows that the algorithm is false alarming on benign objects that appear similar to SMT-parts.

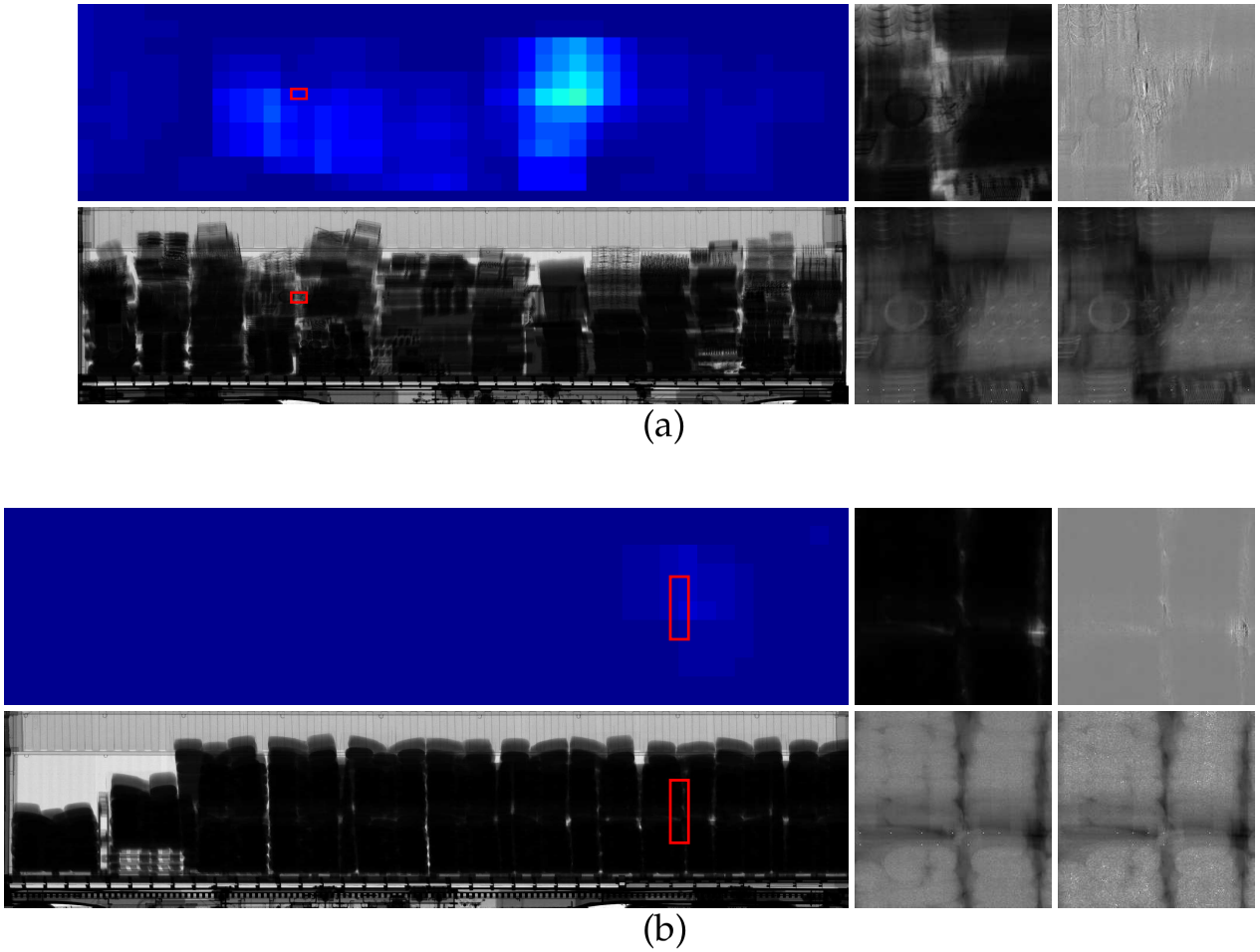


Figure 8.7: Example false negatives. Above each raw image, is a heatmap of the individual window scores. To the right of each, the four patches that are inputs to the network are displayed. The red rectangles indicate the ground truth location of the SMT. The SMT cannot be seen in the network inputs, by eye, and hence detection even for humans is difficult.

8.4.5 *Empty Image Projection*

The EIP validation procedure, as presented in Section 6.5, was performed on the baseline 4-channel dual-energy network. The results are summarised in Table 8.3. The following observations are made:

1. A CNN system trained specifically to detect TIP can indeed learn to exploit TIP artefacts, with slightly better than chance performance in tests (1) and (3);

2. The TIP-trained system has learnt to exploit TIP artefacts, although not as much as the system specifically trained to do so, since it performs slightly better than chance in test (2);

3. In the cases where EIP patches can be detected with good localisation, the EIP patch is projected onto an empty background (Figure 8.8) - in these cases SMTs are straightforward to detect anyway;

4. The system trained on TIP as the threat class and EIP as the benign class, yields very close to chance performance at detecting TIP artefacts in test (5), thus it has been discouraged from learning TIP artefacts;

5. The system trained on TIP as the threat class and EIP as the benign class, yields better performance than a system trained only on TIP.

Test #	Trained on	Tested on	FPR90	FPR95	FPR99
(1)	EIP vs benign	EIP vs benign	89.6	94.5	98.7
(2)	TIP vs benign	EIP vs benign	89.8	94.7	98.9
(3)	EIP vs benign	TIP vs benign	89.6	94.5	98.7
(4)	TIP vs EIP	EIP vs benign	90.0	95.0	98.9
(5)	TIP vs EIP	TIP vs benign	0.46	1.78	18.9
(6)	TIP vs benign	TIP vs benign	0.56	2.38	19.4

Table 8.3: Empty Image Projection test results. Note that for change performance, one would expect FPR90 = 90%, FPR95 = 95%, and FPR99 = 99%.

Therefore, even though a CNN can learn to exploit TIP artefacts the potential performance boost is very small and helps only when SMTs are projected onto low density backgrounds for which SMT detection is straightforward, and the effect can be removed by training on both TIP and EIP. The final point above, that the system trained on TIP versus EIP performs better than a system trained on TIP only, is interesting and unexpected. However, it can potentially be explained by stochastic variation in the trained network or because the artefact-learning parameters in the network have been freed up to

learn to detect actual SMT cues. In any case, it is likely that a system trained and tested on real SMTs would provide similar performance to the system developed on TIP in this work.

Figure 8.8 gives examples of EIP detection in test (1). Examples (a), (b) and (c) show no localisation of the EIP patch, and so the detections are down to chance. Examples (c), (d) and (e) show a localisation of the EIP patch and thus it is likely that the CNN can detect these. It was observed when looking at the test examples, that EIP detections were only well-localised when they were projected onto low density empty container patches, and when projected onto cargo there were no localised detection. Therefore, in the cases where TIP artefacts can be most exploited, SMT detection is straightforward and so the artefacts make very little difference to overall performance metrics.

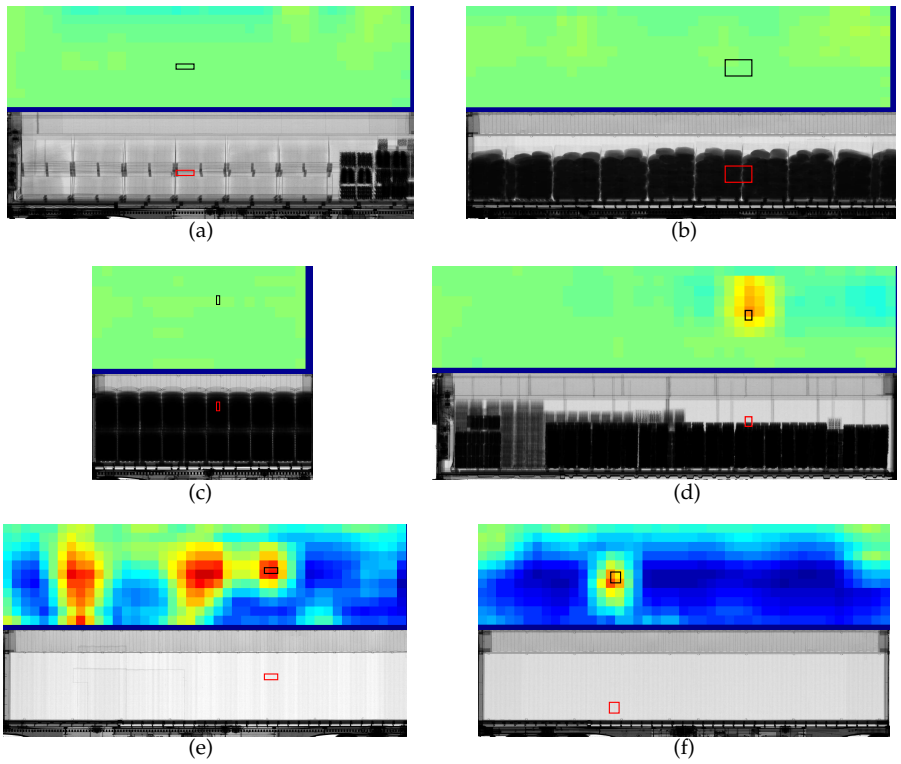


Figure 8.8: Example detection for EIP-trained network on EIP test examples. The red and black rectangles indicate the projection location for EIP. In cases where the projection location coincide with a boost in detection signal, the patches are projected onto empty parts of the container, where SMT detection is trivial even for the worst performing single-energy networks.

In Figure 8.9, the Receiver Operating Characteristic (ROC) curves for tests (1), (3) and (5) are given. Chance performance on the ROC is equivalent to a straight diagonal line through the origin. For test (1), the performance is above chance for low detection rates and this corresponds to EIP onto empty

patches. The performance drops slightly for test (2), and reaches almost chance levels for test (3) when the network trained on TIP versus EIP to discourage it from learning TIP artefacts.

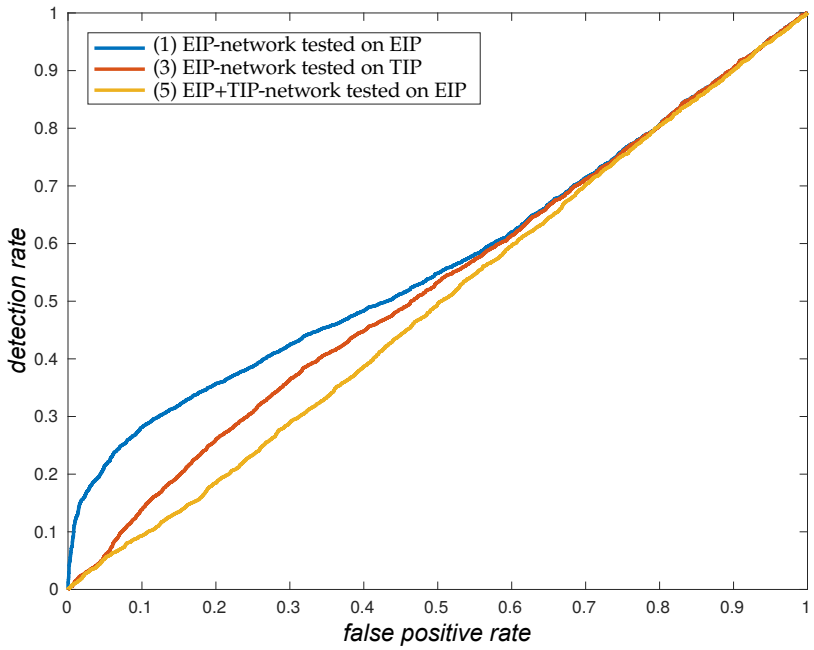


Figure 8.9: ROC curves for EIP tests (1), (3), and (5). A network trained on EIP versus benign achieves above chance performance when tested on EIP or TIP versus benign. This indicates that it is possible to train a network to exploit subtle TIP artefacts. When training is performed on EIP versus TIP, the network achieves chance performance on classifying EIP patches, indicating that it has been forced to ignore (not learn) TIP artefacts.

8.5 DISCUSSION

In this chapter, the use of trained-from-scratch CNNs on complex dual-energy X-ray cargo imagery has been investigated. Dual-energy X-ray imagery can, in theory, reveal information about the types of material in the image. Coupled with spatial information, it is hypothesised that CNNs can implicitly learn to exploit the material information to suppress the number of false alarms in ATD. This work has built upon other co-authored work [10], on ATD of SMTs, by investigating three existing methods of cargo material discrimination [5, 97, 98, 142] and three novel variants, and investigate how they perform when fed to the CNN as separate input channels. Two changes to the training scheme were also proposed, namely, on-the-fly generation and 'spot-the-difference' threat and benign training pairs, complete with data augmentation, on-the-fly. It was found that on-the-fly training provided

a significant boost to test performance. However, that 'spot-the-difference' training offered no statistically significant improvement in performance, but can offer some improvements in training speed.

In the dual-energy experiments, it was found that each dual-energy CNN performed better than its single-energy analogue. This supports the hypothesis that the CNN can implicitly learn how to decode material information from dual-energy images in order to suppress false alarms. Moreover, it was found that a four-channel input, consisting of the sum and difference of the high and low images together with the log-transform of the high and low energy images, yielded the best performance in terms of the number of false positives given detection rates of 95% and 99%. This system was capable of detecting 95% of containers containing a single SMT, while raising 0.34% false positives on benign containers. Overall, the improvement by exploiting dual-energy and the new training scheme, resulted in a 2.5% improvement in AUC, and a 100-fold improvement in the false positive rate when the detection rate is fixed at 90%. Another way of expressing this improvement, is that if the false alarm rate was kept at the level of Jaccard et al. [10] (6%) the system can now detect 98.4% of SMT; an improvement of 8.4% in detection over the previous state-of-the-art.

Furthermore, the EIP test procedure, introduced in Chapter 6, was applied to the baseline dual-energy network. It showed that (i) it is possible to exploit TIP artefacts to give a *slight* performance boost, (ii) but the cases where TIP artefacts can be detected are on empty container backgrounds where ATD is trivial for CNNs, and (iii) the network can be forced not to learn TIP artefacts by projecting TIP artefacts into the benign class during training. Whilst this provides encouraging evidence of high performance in real-life applications, the only concrete evidence would be to properly test the system on a large dataset of real-life concealed SMTs. Such data was not available for this thesis.

9

CONCLUSIONS

This chapter presents an overview of the research conducted, critically reviews the most important novel contributions, and suggests avenues for future research and development.

9.1 RESEARCH SUMMARY

The unfettered movement of goods across borders is an economic necessity. However, high throughput inevitably compromises security, and countries are increasingly exposed to a diverse range of threats from Organised Crime Networks (OCNs), terrorist networks, and even 'legitimate' companies. The origin of this compromise is the reliance on human operators to pore over insurmountable volumes of X-ray images in the search for threats, fraud, or contraband. To the operators, the task is akin to searching for needles in an ever-growing field of haystacks. The field; the vast global container fleet that constitutes the global supply chain. The haystacks; the diverse range of confusing legitimate items that fill a 40 ft shipping container. And ever-growing because the number of containers in the fleet continues to grow each year.

The current security protocol for monitoring the flows of goods in cargo containers is already ill-equipped under the current threat-levels, and this is going to worsen in the future as the amount of trade continues to grow. The security protocol typically consists of three layers:

(i) selection of containers based on a risk analysis, specific intelligence from policing operations, or at random;

(ii) the Non-Intrusive Inspection (NII) of the selected containers to form a radiographic image of the container and its contents which is inspected by a human operator;

(iii) if a threat, fraud, or contraband is found the container is sent for physical inspection which is a long, arduous and expensive process.

The grand goal of the Container Security Initiative (CSI), introduced in the wake of the 9/11 attacks, was to mandate total inspection of the whole container fleet. However, this goal has been far from realised, with some estimating that only 4% of containers are imaged [3] and even fewer images are actually inspected. The other 96% of containers, although deemed lower risk, could and sometimes do contain threats or contraband.

There has been little headway in the use of automated image analysis techniques to reduce or relieve the burden on human operators. Far more attention has been paid to the automated analysis of passenger baggage images in airport security because (i) there is a more imminent perceivable threat from terrorism, (ii) it has, until recently, been better funded by governments, and (iii) the problem of analysing baggage imagery is more tractable. The technical contributions in this thesis have investigated ways to reduce or relieve the burden on human operators through the use of automated image analysis applied to complex transmission X-ray image of cargo containers and vehicles.

Some large-scale X-ray systems are required to move past the container or vehicle to form an image. During this movement the imaging array may wobble which can lead to noise, artefacts or geometric distortions in the image. These wobble effects can reduce the ability of the operator to quickly identify threats since they reduce the quality of images and can also effect the precision of material discrimination. In this thesis, a method was introduced to measure and correct for wobble noise and artefacts. The method relied on the use of Beam Position Detectors (BPDs); imaging detectors rotated by 90° so that they can measure the profile of the X-ray beam across its width. This beam profile encodes information about the position of the beam-centre. If tracked through the scan the amount of wobble can be estimated across the image. However, estimation is not simple since the beam is attenuated by arbitrary objects in the scene. An algorithm was developed for estimating wobble based on BPD measurements. To estimate the beam position at a particular instance, a Bayesian fusion was performed of (i) an instantaneous estimate from that scanning instance, and (ii) a prior estimate based on previous beam position estimates. The Bayesian fusion meant that if the instantaneous estimate is very confident then it contributed more to the fused estimate; if it was uncertain then the prior estimate was relied upon. This enabled superior beam position estimates particularly in difficult cases. The instantaneous estimation was based on a Random Regression Forest (RRF)

operating on the estimated beam profile before it interacts with the scene. The prior estimate was obtained by fitting an Auto-Regression (AR) model to a training set.

With these wobble estimates the image could be corrected for wobble. The correction was derived by considering a model of image formation in the presence of a wobbling detector. It was found that several other imperfections in the scanning apparatus complicated wobble correction, and so methods to estimate these and their corrections were also proposed. The correction method was tested on (i) images of an air-only scene, and (ii) images of scenes containing objects of increasing complexity. The former was useful for quantitatively evaluating the improvement in image quality by the wobble correction method, and the later was useful for qualitative evaluation. It was found that the wobble correction method yielded an 87% improvement in image noise due to wobble.

A problem in training and evaluating Machine Learning (ML) based systems for Automated Threat Detection (ATD) in security, is the scarcity of data. For machines the *data problem* for threat items means that it is difficult to train sophisticated ML-based systems without overfitting, and difficult to evaluate performance due to the *accuracy paradox*. Potential solutions include:

(i) the collection of large amounts of staged threats, which is particularly expensive for cargo;

(ii) treating ATD as a one-class anomaly detection problem, but this is difficult and is likely to result in lots of false positives unless the benign dataset is sufficiently large;

(iii) transfer-learn systems trained on other modalities where data is plentiful, however this is unlikely to perform as well as a bespoke-trained network.

Another solution is to learn from human operators. It has been shown that the rarer a threat item, the lower the performance of a human operator at detecting that threat item. In aviation baggage, the remedy was to introduce Threat Image Projection (TIP). TIP works by injecting a Fictional Threat Image (FTI) into baggage images. This (i) can be used in Computer-Based Training (CBT) of operators, (ii) exposes the operator to more threat items to reduce to combat the prevalence effect, and (iii) allows the performance of operators

to be evaluated. Motivated by an analogy between humans and machines, a TIP framework was developed for cargo.

The TIP framework proposes a number of variations that can be applied to the threat before projection, and can be used as data augmentation to boost the number of unique examples when training ML-based systems. Such methods have been used in the mainstream computer vision community to improve performance and prevent overfitting of Convolutional Neural Networks (CNNs). The other advantage of TIP is that, when testing systems, one can control a number of conditions in the testing to build a better picture of system performance. The method for extracting and projecting threat items was validated experimentally. TIP images were compared qualitatively and quantitatively to real threat images, and it was found that the two were indistinguishable. However, still concerned that a system can learn to exploit potential TIP artefacts, such as halos and subtle changes in noise distribution, to artificially boost performance, a method coined Empty Image Projection (EIP) was introduced. The aim of EIP is to project just the TIP artefacts and not the threat itself. Using a series of tests involving EIP, one can test whether systems are capable of achieving improved performance on TIP imagery over real imagery. Both TIP and EIP were utilised in the final two technical contributions.

A large proportion, 20%, of the global container fleet are declared-as-empty. These containers are convenient and cheap for criminals to smuggle contraband either in legitimate or false container partitions, or unhidden in rip-on/rip-off attacks. False declared-as-empty containers can also be used to avoid tariffs and sanctions, or can be a health and safety hazard if placed in an empty-only container stack at ports. An automated system, capable of verifying whether declared-as-empty containers are truly empty, can therefore be used used to inspect 20% of all containers. It does not need to detect specific threats; any foreign object in the container is an indication of illegal behaviour or error.

Indeed, false declared-as-non-empty containers can also pose safety hazards, but it is difficult to think of criminal reasons to do so, and it is not reported in the literature. Perhaps errors or theft can lead to false declared-as-non-empty containers.

Automated Empty Container Verification (ECV) is complicated by the diverse range of appearances of declared-as-empty containers, and because small amounts of smuggled contraband can appear similar to container damage, packaging, or detritus. A system for automated ECV that operates on full-sized images was proposed. The system is based on the Random Forest (RF) classification of local geometric features together with the coordinates of the analysis window. It is hypothesised that the addition of the

window coordinates as a feature, allows the RF to implicitly learn the range of appearances at different locations within the container and avoids the need to segment the container into different regions and learn a separate classifier for each region. The system is trained using the TIP framework to generate examples of synthetic non-empty containers. It is tested on both real non-empty containers from the Stream-of-Commerce (SoC) and TIP imagery. The test on TIP allows (i) for non-empty containers much more difficult than those found in the SoC to be generated, and (ii) for performance to be assessed as a function of difficulty (volume and density of the load) and location. The system is able to detect 99.3% of non-empty containers in the SoC whilst raising 0.7% false positives, and is capable of detecting >93% of loads similar in size and density to 1.25 kg whilst raising 1% false alarms.

The final contribution of this thesis was a system for ATD based on complex dual-energy X-ray cargo images. Several methods for exploiting dual-energy measurements in CNNs were explored and compared. Three of these methods were based on existing methods for material discrimination in the literature, and three were novel variants. The aim was to provide a CNN with material information that it could use to implicitly learn about materials in order to suppress false alarms in ATD. For material discrimination methods used alone as input channels, it was found that the best performing method was a novel variant ($\{r, \theta\}$ method). Combining two methods, together as input channels, lead to a 100-fold reduction in false alarms over other co-authored work not presented in this thesis [10]. Moreover, the EIP framework was used to show that the system can obtain a very small performance boost due to TIP artefacts. However, this effect could be removed by training the system on both TIP imagery and EIP imagery so as to encourage the system not to learn cues from TIP artefacts. To best knowledge, this is the first time that dual-energy cargo imagery has been used for the purposes of automated *image understanding*, and the first time CNNs have been applied to raw dual-energy measurements, in any field.

9.2 REVIEW OF CONTRIBUTIONS

The most pertinent novel contributions of this work are summarised as follows.

LITERATURE REVIEW The first review of the field of Automated X-ray Image Analysis for Cargo Security. Due to the relevant infancy of the field, no previous review had previously been published in the literature. The review covered all aspects of image analysis, including *image pre-processing* and *image understanding*. The published version of the review [7] is longer than the version presented in this thesis, with more analysis on the future direction and promise of the field.

WOBBLE MEASUREMENT A novel method of measuring detector wobble in large-scale transmission radiography was developed. The method relied on rotating a few imaging detectors by 90°, to form BPDs which measure the profile of the X-ray beam across its width. During a scan the BPDs are obscured by arbitrary objects in the scene, thus complicating wobble estimation. The proposed solution used a RRF model to obtain an instantaneous wobble estimate, and an AR model to determine an estimate based on previous estimates. The RRF and AR estimates were then Bayesian fused to form a superior estimate of wobble.

WOBBLE CORRECTION A novel method for correcting image intensity variation due to detector wobble was presented. The method was derived by considering a model of image formation in the presence of a wobbling detector. Wobble correction also relied on the estimation of misplacements and rotations of individual imaging detectors that can complicate wobble correction. The correction method, when combined with the measurement method, was able to correct for 87% of image intensity variation due to wobble. This work, and wobble estimation, extends the work of Rogers et al. [11], which achieved a wobble correction of 70%.

TIP, DATA AUGMENTATION, AND EIP A method for TIP in cargo was proposed and validated experimentally. The TIP method relies on the Beer-Lambert law, and the experimental validation shows that there is no qualitative or quantitative difference between a real threat image and its equivalent TIP image. Methods of injecting natural variation into TIP imagery were proposed as a form of data augmentation, useful for training ML-based algorithms. Moreover, due to concerns that the ML systems in this thesis could learn subtle artefactual cues present in TIP imagery, an EIP method was proposed and applied to improve confidence that this was not occurring. The

aim of EIP is to project just the TIP artefacts and not the actual threat. According to the literature, this was the first proposal of using TIP to train ATD algorithms, and the first time an EIP was proposed. Thus, there was no prior state-of-the-art in cargo to compare to, but the TIP method was used in the next two chapters which do achieve state-of-the-art performance.

EMPTY CONTAINER VERIFICATION The first ML-based method for ECV was presented. The method works by RF classification of individual image patches. The features included fixed geometric features and the coordinates of the patch within the image. The use of window coordinates as a feature allowed the RF model to learn the location-specific range of appearances of an empty container and overcame the need to segment the image and apply a separate classifier to each image region. The method was trained on TIP-generated non-empty examples and real empty examples from the SoC. The proposed ECV method outperformed a previous method reported in the literature when tested on the SoC, and was also tested for very small TIP-generated adversarial loads representative of smuggled cocaine. The system was also validated using EIP. The system achieved a 99.3% detection rate and 0.7% false positive rate for real SoC images, and extends the work of Orphan et al. [66], which achieved an accuracy of 97.2% with 0.4% false positive rate. The evaluation was also extended to include loads resembling small amounts of smuggled cocaine, and the system was able to detect 93% of 1.25 kg with a 1% false alarm rate.

DUAL-ENERGY AUTOMATED THREAT DETECTION A method for detecting so-called Small Metallic Threats (SMTs) in complex dual-energy cargo imagery was presented. According to the literature, it was the first time in cargo that ATD has explicitly operated on images measured at different energies. The motivation was that the algorithm can learn to suppress false positives by using material information derived from the dual-energy measurement. The method employed a trained-from-scratch CNN, trained on TIP imagery with data augmentation, and significantly improved on the prior work of Jaccard et al. [10], decreasing the false positve rate from 6% to 0.08%, for 90% detection. The system was also validated using EIP, which showed that there was no evidence that the system was exploiting potential cues from TIP artefacts to boost performance. Although this can still not be completely ruled out, unless a test is performed on real threat imagery.

9.3 CRITICISMS AND FUTURE WORK

As with all research, whilst this thesis has provided several important contributions to automated cargo image analysis, it has opened up several more avenues of research. It is recommended that these are addressed in future work.

9.3.0.1 *Wobble*

In this thesis, the effects of wobble noise and artefacts manifesting as image intensity variation were investigated. This wobble is due to the movement of the detector boom from side to side like a swinging pendulum. However, very severe wobble can make the detector boom move up and down relative to the source. In such cases, wobble manifests as a geometric distortion in images, where straight horizontal lines in the scene become wobbly lines in the image. This can also affect the ability of the operator or algorithm to detect threats. So correction of geometric distortions due to wobble is an area for future research.

In addition, the wobble measurement method struggled when large and extremely dense objects occluding the BPD during the scan leading to a low Signal-to-Noise Ratio (SNR) and poor instantaneous estimates. One can also imagine scenarios where the BPDs are completely occluded such that there is no signal reaching it at all. In these cases, extra information is required to measure wobble. A possible way of providing this information is through the use of extra measurement devices, such as accelerometers, placed on the detector boom to improve the prior beam position estimate in cases where the instantaneous estimate fails. Moreover, wobble estimation could also be improved by using more BPDs or even a 2D imaging array. A 2D imaging array could be expensive, but would also result in better quality images since one would obtain more samples of each point in the scene.

9.3.0.2 *Real test data*

Although every effort was made to validate the use of TIP in training and testing ATD, through experimental validation and the proposed EIP method, the only truly conclusive test is on real data of smuggled threats. Such data was not available during this thesis, and is very difficult to obtain for cargo. However, in the future real data should be obtained and used for testing purposes. Until cargo data is available, a simple and meaningful approach

Varied baggage threat data is also easier to stage, than loading and unloading a 40 ft container.

is to switch the system over to airport baggage data. Baggage data is much easier to obtain, and the methods used here are equally applicable to baggage. The system would be trained in exactly the same way using TIP, but it would be tested on real baggage threat examples, to test whether the performance on TIP matches the performance on real data. If it does not, then one can discover why not, and use this to improve the TIP used in training.

9.3.0.3 *Positional information*

It is likely that better ECV performance can be obtained using a dual-energy CNN approach similar to the method used in ATD. Perhaps, it would also be beneficial to give the CNN positional information, so that the CNN can learn the expected range of appearances across the container, as it was found with the RF. This is a potential avenue for future research. It is also likely that for ATD, positional information can also boost the performance. For example, a system performing threat detection based on oriented Basic Image Features (oBIFs), or similar Bag-of-Words (BoW) approaches, would benefit from positional information, since relatively rare container structures such as metal roof bows, can appear similar to threats. However, it is unclear whether such information would help much for CNN approaches because performance is already very high. It was noted however, that the CNN-based ATD performed worst in the bottom corner of the container. This region tends to be dense making ATD more difficult than other parts of the container, however it is not as dense as some of the loads in which the system was still capable of detecting threats. Perhaps this could be improved by the use of positional information, or it could be because threats were less likely to be placed in the bottom corner during training and therefore the CNN was exposed to fewer of such examples.

9.3.0.4 *Anomaly detection*

When human operators inspect cargo images, they use two methods: (i) threat detection, and (ii) anomaly detection. Threat detection looks for the tell-tale silhouette of the threat in the image. As the X-ray image becomes more cluttered and the threat is placed in more complex or dense surroundings, detection of this tell-tale silhouette becomes more difficult. In these cases, operators rely on spotting anomalies that may be indicative of a threat. Criminals who have knowledge of X-ray screening may try to shield a threat using dense materials or complicated, confusing clutter. The ATD algorithm

developed in Chapter 8 detects threat silhouettes, and works even in cases where there is dense or confusing clutter. However, in cases where the threat is completely shielded such that there is no threat information left, detection fails. In such cases, an Automated Anomaly Detection (AAD) algorithm could be used to assess the suspiciousness of the shielded regions that may be indicative of smuggling activity and flag them up to border staff. For example, a dense region may be suspicious relative to the other content or due to its positioning. AAD would also be useful for other reasons, including but not limited to: (i) detection of false partitions; (ii) detection of emergent threats which an ATD algorithm has not been trained on; (iii) detection of unusual weight distributions that can be indicative of smuggling.

AAD can also overcome the *data problem*, because such systems only need to be trained on benign data, which is easier to obtain. However, it is likely that an AAD system will generate high numbers of false positives unless a very large amount of benign data is used. AAD is a complicated problem, but systems able to obtain performance similar to or better than human operators, could provide a significant boost to cargo security, and so remains an important avenue of future research.

9.3.0.5 *Generalisability*

There are questions on how well the ECV and ATD algorithms generalise to other types of cargo scanners with slight differences in image properties. It would be worthwhile assessing the performance of the algorithms developed for the Rapiscan Eagle® R60, directly on other scanners. If the performance does not generalise well, then can it be improved by adding more data augmentation or learning a function that maps image appearance between scanners?

9.4 CONCURRENT WORK

Most research that was published concurrent to this thesis were summarised in the literature review. However, there has been some recent work by the ACXIS project [178, 179], which has not been discussed in this thesis. The work focusses on providing tools for *assisted inspection*. These tools include:

Automated Comparison of X-ray Images for Cargo Screening (AXCIS).

- *Image standardisation:* The project has developed a tool for standardising X-ray images captured from different cargo scanners, with the aim of developing algorithms capable of ATD across a range of devices.

This includes adjustment of the image contrast, and geometric transformations to standardise the acquisition geometry [178].

- *TIP:* The project has developed tools for synthesising X-ray images from 3D Computer Aided Design (CAD) models that can then be projected into real X-ray images. In addition, they have developed a TIP method based on the Beer-Lambert law and similar to the method proposed in Chapter 6. They call this method 'image blending', and give a visual comparison between TIP and a real threat image [178]. However, they are not compared quantitatively, but the authors state that 'According to expert judgement by customs officers and X-ray imaging experts, the [method] produces realistic results' [178]. It is unclear whether the TIP imagery is used in, or proposed for, direct training of algorithms.

- *Cigarette detection:* Few technical details of the algorithms are given. However, images are pre-processed with Block Matching 3D (BM3D), as it was found to perform better than several other approaches including: Gaussian filtering, bilateral filtering, anisotropic filtering, and Non-Local Means (NLM) [178]. No details of the detection algorithms are given, but Visser et al. [178] show an example of detection where a container is half-filled with cigarettes, and they state '[cigarettes] are typically transported in large quantities and appear in X-ray scans as homogeneous regions with specific textures due to the common way of packaging layers of cartons into boxes'. This implies that the algorithm is based on texture recognition and is designed for detecting only very large quantities of cigarettes.

- *Image comparison:* for the purposed of aiding narcotics detection, the project has developed a tool for comparing an image with a reference benign image. In this, the user will select an analysis Region-Of-Interest (ROI), and a similar ROI is found in a reference database by comparing 'intensity and structural features' [178]. The two ROIs are registered by computing Speeded-Up Robust Features (SURF), and filtering pairs of keypoints using a nearest neighbour approach and RANdom SAmple Consensus (RANSAC) [178]. Based on these keypoints, an affine or homographic transformation is determined. With the images registered, the images are then compared by creating a distance map using adaptive thresholding and morphological filtering.

This difference map can be used by operators to identify anomalies relative to the reference image [178, 179].

- *Other functions:* The project has also developed tools for counting the recurrence of a manually selected ROI [178, 179], and for weight estimation. However, few technical details are given.

Previously, we defined a number of historical stages in the development computer vision algorithms. These stages, included:

(i) algorithms completely hand-crafted by experimentation and intuition;

(ii) features hand-crafted based on intuition and experimentation, with the features classified using ML techniques; and

(iii) the features and their classification learnt directly from data.

The ACXIS project demonstrates that there is a concerted effort in Europe to improve inspection procedure using automated image analysis. However, it also demonstrates that this effort is still stuck in stages (i) and (ii). But there are indications that future research will include the used of CNNs [178] bringing the project to stage (iii).

9.5 FUTURE OF THE FIELD

It is not difficult for one to imagine a future where the global supply chain is fully automated. Signs of it are already beginning to emerge with delivery companies experimenting with the use of drones [180] for parcel delivery and driverless delivery vehicles [181], and the recent opening of a fully automated London Gateway Port [55]. The need for automation is driven by the need to cope with the increasing throughput demands. And with increasing levels of automation in logistics, it is likely that the levels of throughput will grow even faster than they do today.

On the screening side, there are now commercial scanners that can scan containers in less than a second, thus preserving high throughput. The main limitation in terms of both security and throughput, is the use of human operators, who have many undesirable features, including: (i) their susceptibility to bribery and blackmail, (ii) their inconsistent performance in terms of accuracy and time, (iii) their need for holidays and breaks due to illness, (iv) their expensive training, and (v) the difficulty of scaling them.

This thesis has proposed methods to assist human operators, which is essential as the volumes of trade continue to grow. With future developments in artificial intelligence, it is not inconceivable that human operators will eventually be replaced altogether, or freed up to perform the more difficult tasks such as physical inspection and intelligence gathering.

Innovations in automated inspection may drive innovations in hardware. For example, if algorithms are capable of outperforming humans, then the scanning hardware can be tweaked to allow higher scanning speeds. Scanning at faster rates leads to lower resolution and to a lower SNR, but algorithms could still obtain high performance by learning from vast numbers of examples. In addition, there may be the development of novel scanner designs that are optimised for algorithms rather than humans. In particular, scanners could become driven by artificial intelligence and adapt scanning to current risk levels, or intelligently adjust scanning to focus on regions that are suspicious or difficult to penetrate.

With increased automated inspection, is it possible that new attack vectors will emerge? Criminals, as intelligent and rational actors, will invent and trial new ways of defeating security systems. One clear example would be increased use of cyber-attacks on port security systems. Such attacks have already been documented [182], but perhaps they will become more commonplace as the *modus operandi* of using human agents becomes more difficult. Such cyber-attacks could have several goals:

(i) as denial of service, to prevent algorithms from accessing resources such as manifest information or cargo images;

(ii) as denial of service, to force ports to switch back to human image inspection only;

(iii) as a means of poisoning data with false data, so that systems, particularly those using online learning, become corrupted;

(iv) as a means of spoofing incorrect algorithm decisions to cause distrust of new systems;

(v) as a means of algorithm theft, so that algorithms can be probed by adversarial artificial intelligence to generate new smuggling methods capable of defeating automated systems.

As with all new security technologies, heed of these potential attack vectors is needed if the technology is to be a success.

Nevertheless, it is clear that future algorithmic developments will play an important role in lowering the burden on human operators in their search for needles in the ever-growing field of many haystacks.

BIBLIOGRAPHY

[1] The World Bank. 'Container port traffic (TEU: 20 foot equivalent units)'. url: http://data.worldbank.org/indicator/IS.SHP.GOOD.TU/countries. Accessed: 14-06-2016 (cited on pp. 1, 10).

[2] L. Tavasszy et al. (2011). 'A strategic network choice model for global container flows: specification, estimation and application'. *Journal of Transport Geography* **19**.6, pp. 1163 –1172 (cited on p. 1).

[3] House Hearing, 112 Congress. 'Balancing maritime security and trade facilitation: protecting our ports, increasing commerce and securing the supply chain'. url: https://www.gpo.gov/fdsys/pkg/CHRG-112hhrg76511/html/CHRG-112hhrg76511.htm. Accessed: 01-01-2017 (cited on pp. 1, 19, 164).

[4] A. Chalmers (2007). 'Automatic high throughput empty ISO container verification'. In: *Proc. SPIE*. **6540**, pp. 1–4 (cited on pp. 2, 57, 58).

[5] S. Ogorodnikov and V. Petrunin (2002). 'Processing of interlaced images in 4–10 MeV dual energy customs system for material recognition'. *Physical Review Special Topics – Accelerator and Beams* **5**.10, pp. 67–77 (cited on pp. 2, 49–53, 55, 56, 72, 95, 144–146, 161).

[6] J. Zhang et al. (2014). 'Joint Shape and Texture Based X-ray Cargo Image Classification'. In: *Proc. IEEE Computer Vision and Pattern Recognition Workshop*, pp. 266 –273 (cited on pp. 2, 44, 57, 59, 60).

[7] T. W. Rogers et al. (2016). 'Automated X-ray image analysis for cargo security: Critical review and future promise'. *Journal of X-Ray Science and Technology* **25**.1, pp. 33–56 (cited on pp. 3, 5, 6, 168).

[8] BBC. 'As it happened: Mumbai attacks 27 Nov'. url: http://news.bbc.co.uk/1/hi/world/south_asia/7752003.stm. Accessed: 10-09-2017 (cited on pp. 4, 18).

[9] M. Roomi and R Rajashankari (2012). 'Detection of Concealed Weapons in X-ray Images using Fuzzy K-NN.' *International Journal of Computer Science, Engineering and Information Technology* **2**.2, pp. 187–196 (cited on p. 4).

[10] N. Jaccard et al. (2016). 'Automated detection of smuggled high-risk security threats using Deep Learning'. In: *Proc. IET Imaging for Crime*

Detection and Prevention, pp. 11–15 (cited on pp. 5, 70, 141, 147, 148, 150–153, 161, 162, 167, 169).

[11] T. W. Rogers et al. (2014). 'Reduction of Wobble Artefacts in Images from Mobile Transmission X-ray Vehicle Scanners'. In: *Proc. IEEE Imaging Systems and Techniques*, pp. 356–360 (cited on pp. 5, 77, 84, 89, 91, 95, 106, 115, 168).

[12] T. W. Rogers et al. (2016). 'Measuring and correcting wobble in large-scale transmission radiography'. *Journal of X-Ray Science and Technology* **25**.1, pp. 55–77 (cited on pp. 5, 6).

[13] T. W. Rogers et al. (2016). 'Threat Image Projection (TIP) into X-ray images of cargo containers for training humans and machines'. In: *Proc. IEEE International Carnahan Conference on Security Technology*, pp. 1–7 (cited on pp. 5, 6).

[14] T. W. Rogers et al. (2015). 'Detection of cargo container loads from X-ray images'. In: *Proc. IET Intelligent Signal Processing*, pp. 1–6 (cited on pp. 5, 6).

[15] T. W. Rogers, N. Jaccard and L. D. Griffin (2017). 'A deep learning framework for the automated inspection of complex dual-energy x-ray cargo imagery'. In: *Proc. SPIE*. **10187**, pp. 1–12 (cited on p. 5).

[16] M Caldwell et al. (2017). 'Transferring x-ray based automated threat detection between scanners with different energies and resolutions'. In: *Proc. SPIE*. **10441**, p. 1 (cited on p. 5).

[17] N. Jaccard et al. (2016). 'Tackling the X-ray cargo inspection challenge using machine learning'. In: *Proc. SPIE*. **9847**, pp. 1–13 (cited on pp. 5, 100).

[18] — (2015). 'Using deep learning on X-ray images to detect threats'. In: *Proc. Cranfield Defence and Security Doctoral Symposium*, pp. 1–12 (cited on pp. 6, 124).

[19] — (2016). 'Detection of concealed cars in complex cargo X-ray imagery using deep learning'. *Journal of X-ray Science and Technology* **25**.3, pp. 323–339 (cited on pp. 6, 67, 68, 116).

[20] N. Jaccard, T. W. Rogers and L. D. Griffin (2014). 'Automated detection of cars in transmission X-ray images of freight containers'. In: *Proc. IEEE Advanced Video and Signal Based Surveillance*, pp. 387–392 (cited on pp. 6, 60, 67, 116, 124).

[21] J. T. A. Andrews et al. (2017). 'Representation-learning for anomaly detection in complex x-ray cargo imagery'. In: *Proc. SPIE*. **10187**, pp. 1–11 (cited on pp. 6, 68).

[22] J. T. A. Andrews et al. (2016). 'Anomaly Detection for Security Imaging'. In: *Proc. Cranfield Defence and Security Doctoral Symposium*, pp. 1–14 (cited on p. 6).

[23] The Economist. 'Machines are learning to find concealed weapons in X-ray scans'. url: http://www.economist.com/news/science-and-technology/21711016-artificial-intelligence-moves-security-scanning-machines-are-learning-find. Accessed: 19-12-2016 (cited on p. 6).

[24] H. H. Willis and D. S. Ortiz (2004). 'Evaluating the Security of the Global Containerized Supply Chain'. RAND Corporation (cited on pp. 9, 10).

[25] National Research Council (US). Committee for a Study of the Effects of Regulatory Reform on Technological Innovation in Marine Container Shipping (1992). 'Intermodal marine container transportation: impediments and opportunities'. (cited on p. 9).

[26] M. Richardson (2004). 'A Time Bomb for Global Trade: Maritime-related Terrorism in an Age of Weapons of Mass Destruction'. *Maritime Studies* **2004**.134, pp. 1–8 (cited on p. 9).

[27] E. Iritany and M. Dickerson (2002). 'Calculating cost of West Coast dock strike is a tough act'. *Los Angeles Times* **26**. (cited on p. 10).

[28] S. S. Cohen (2002). 'Economic impact of a West Coast dock shutdown'. *University of California at Berkeley*, p. 1 (cited on p. 10).

[29] Y. Liu, B. D. Sowerby and J. R. Tickner (2008). 'Comparison of neutron and high-energy X-ray dual-beam radiography for air cargo inspection.' *Applied Radiation and Isotopes* **66**.4, pp. 463–473 (cited on pp. 10, 20, 72).

[30] J. Romero (2003). 'Prevention of maritime terrorism: the Container Security Initiative'. *Chicago Journal of Inertnational Law* **4**., pp. 597–605 (cited on p. 10).

[31] U.S. Customs and Border Protection. 'Container Security Initiative In Summary'. url: www.cbp.gov/sites/default/files/documents/csi_brochure_2011_3.pdf. Accessed: 14-06-2016 (cited on p. 10).

[32] HM Government. 'Serious and Organised Crime Strategy'. url: https://www.gov.uk/government/uploads/system/uploads/attachment_

data/file/248645/Serious_and_Organised_Crime_Strategy.pdf. Accessed: 07-12-2016 (cited on p. 11).

[33] D. B. Cornish and R. V. Clarke (2014). 'The reasoning criminal: Rational choice perspectives on offending'. Transaction Publishers. ISBN: 1412852757 (cited on p. 11).

[34] M. Dubbey. 'Former Head of Drugs Intelligence for UK SOCA (NCA) and HMRC'. Private Communication (cited on pp. 11, 15, 18).

[35] Y. Y. Haimes (1981). 'Hierarchical holographic modeling'. *IEEE Transactions on Systems, Man, and Cybernetics* **11**.9, pp. 606–617 (cited on p. 11).

[36] OECD. 'Magnitude of counterfeiting and piracy of tangile products: an update'. url: https://www.oecd.org/sti/ind/44088872.pdf. Accessed: 08-05-2017 (cited on p. 11).

[37] European Commission (2002). 'Types of Fraud and Trends'. *Good Practice Guide for Sea Container Control* **Ch. 4**. (cited on pp. 11, 17, 23).

[38] M. McNicholas (2016). 'Maritime security: an introduction'. Butterworth Heinemann. ISBN: 9780128036723 (cited on pp. 13, 14).

[39] European Commission (2002). 'Concealment Methods'. *Good Practice Guide for Sea Container Control* **Ch. 6**. (cited on pp. 13–15, 111, 112, 138).

[40] — (2002). 'System and Method of Examination'. *Good Practice Guide for Sea Container Control* **Ch. 5**. (cited on pp. 14, 19, 22, 23).

[41] K. M. Finklea. 'Southwest Border Violence: Issues in Identifying and Measuring Spillover Violence'. url: https://www.fas.org/sgp/crs/homesec/R41075.pdf. Accessed: 17-11-2016 (cited on p. 17).

[42] L. I. Shelley and J. T. Picarelli (2002). 'Methods not motives: Implications of the convergence of international organized crime and terrorism'. *Police Practice and Research* **3**.4, pp. 305–318 (cited on pp. 17, 18).

[43] G. E. Curtis and T. Karacan (2002). 'The nexus among terrorists, narcotics traffickers, weapons proliferators, and organized crime networks in Western Europe'. In: *The Library of Congress* (cited on pp. 18, 19).

[44] BBC. 'Paris attacks: What happened on the night'. url: http://www.bbc.co.uk/news/world-europe-34818994. Accessed: 08-10-2017 (cited on p. 18).

[45] — 'Tunisia attack on Sousse beach kills 39'. url: http://www.bbc.co.uk/news/world-africa-33287978. Accessed: 08-10-2017 (cited on p. 18).

[46] N. Bajekal and V. Walt. 'How Europe's Terrorists Get Their Guns'. url: http://time.com/how-europes-terrorists-get-their-guns/. Accessed: 17-11-2016 (cited on p. 18).

[47] M. Felson and R. V. Clarke (1998). 'Opportunity makes the thief'. *Police research series* **98**. (cited on p. 18).

[48] P. Mayhew et al. (1976). 'Crime as opportunity. Home Office research study no. 34'. *London, Home Office* (cited on p. 18).

[49] J. Otis (2014). 'The FARC and Colombia's Illegal Drug Trade'. *The Wilson Center: Latin American Program* (cited on p. 18).

[50] Z. Zhu, Y.-C. Hu and L. Zhao (2010). 'Gamma/X-ray linear pushbroom stereo for 3D cargo inspection'. *Machine Vision and Applications* **21**.4, pp. 413–425 (cited on p. 20).

[51] N. Calvert, E. J. Morton and R. D. Speller (2013). 'Preliminary Monte Carlo simulations of linear accelerators in Time-of-Flight Compton Scatter imaging for cargo security'. *Crime Science* **2**.1, pp. 1–12 (cited on pp. 20, 40, 72).

[52] C. Morris et al. (2008). 'Tomographic Imaging with Cosmic Ray Muons'. *Science & Global Security* **16**.1-2, pp. 37–53 (cited on p. 21).

[53] B. G. Brogdon, H. Vogel and J. D. McDowell (2003). 'A radiologic atlas of abuse, torture, terrorism, and inflicted trauma'. CRC Press. ISBN: 0849315336 (cited on pp. 21, 27).

[54] Evergreen Shipping Agency. 'Customs Scan and Physical Inspection'. url: https://goo.gl/cJ87Mp. Accessed: 10-09-2017 (cited on p. 22).

[55] The Guardian. 'Inside the London megaport you didn't know existed'. url: https://www.theguardian.com/technology/2016/dec/14/amazon-claims-first-successful-prime-air-drone-delivery. Accessed: 31-12-2016 (cited on pp. 23, 174).

[56] Her Majesty's Government. 'National Security Strategy and Strategic Defence and Security Review 2015'. url: www.gov.uk/government/uploads/system/uploads/attachment_data/file/555607/2015_Strategic_Defence_and_Security_Review.pdf. Accessed: 31-12-2016 (cited on p. 23).

[57] D. Customs. 'Scan system operator en Coördinator DCL Maasvlakte'. Private Communication (cited on p. 24).

[58] S. Teerapittayanon, B. McDanel and H Kung (2016). 'Branchynet: Fast inference via early exiting from deep neural networks'. In: *Proc. Pattern Recognition* (cited on p. 25).

[59] S. C. Bushong (2017). 'Radiologic Science for Technologists: Physics, Biology, and Protection'. Elsevier. ISBN: 0323081355 (cited on p. 27).

[60] B. G. Brogdon (1998). 'Forensic Radiology'. CRC Press. ISBN: 9781420075625 (cited on p. 27).

[61] D. Chapman et al. (1997). 'Diffraction enhanced x-ray imaging'. *Physics in medicine and biology* **42**.11, p. 2015 (cited on p. 28).

[62] M. J. Berger et al. (1998). 'XCOM: Photon Cross Sections Database'. url: `http://www.nist.gov/pml/data/xcom/`. Accessed: 15-06-2016 (cited on pp. 31, 146, 147).

[63] H Hirayama (2000). 'Lecture note on photon interactions and cross sections'. *KEK, High Energy Accelerator Research Organization Ibaraki, Japan* (cited on pp. 32, 33).

[64] G. Chen (2005). 'Understanding X-ray cargo imaging'. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* **241**.1, pp. 810–815 (cited on pp. 33, 39, 44).

[65] W. Reed and E Haines (2008). 'Throughput performance factors in X-ray cargo screening systems'. *Port Technology International* **39**. (cited on p. 35).

[66] V. J. Orphan et al. (2005). 'Advanced γ ray technology for scanning cargo containers'. *Applied radiation and Isotopes* **63**.5, pp. 723–732 (cited on pp. 36, 57, 58, 66, 72, 127, 138, 169).

[67] European Commission (2002). 'Container Specifications'. *Good Practice Guide for Sea Container Control* **Ch. 3**. (cited on p. 39).

[68] Y. Zheng and A. Elmaghraby (2013). 'A vehicle threat detection system using correlation analysis and synthesized X-ray images'. In: *Proc. SPIE*. **8709**, pp. 1–10 (cited on pp. 41, 42, 44, 57, 59).

[69] H Vogel (2007). 'Vehicles, containers, railway wagons'. *European Journal of Radiology* **63**.2, pp. 254–262 (cited on p. 42).

[70] A. Chalmers (2007). 'Cargo identification algorithms facilitating unmanned/unattended inspection at high throughput terminals'. In: *Proc. SPIE*. **6736**, pp. 1–6 (cited on pp. 42, 57, 58).

[71] C. Solomon and T. Breckon (2010). 'Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab'. Wiley-Blackwell. ISBN: 0470844736 (cited on p. 43).

[72] S. Michel et al. (2014). 'Increasing X-ray image interpretation competency of cargo security screeners'. *International Journal of Industrial Ergonomics* **44**.4, pp. 551–560 (cited on pp. 43, 47).

[73] B. Klock (2005). 'Test and evaluation report for X-ray detection of threats using different X-ray functions'. In: *Proc. IEEE International Carnahan Conference on Security Technology*, pp. 182–184 (cited on p. 43).

[74] K. Fu et al. (2010). 'The application of wavelet denoising in material discrimination system'. In: *Proc. SPIE*. **7538**, pp. 1–12 (cited on pp. 44, 49, 52, 53, 146).

[75] M. D. Silver, A. Sen and S. Oishi (2000). 'Determination and correction of the wobble of a C-arm gantry'. In: *Proc. SPIE*. **3979**, pp. 1459–1468 (cited on p. 45).

[76] R Fahrig and D. Holdsworth (2000). 'Three-dimensional computed tomographic reconstruction using a C-arm mounted XRII: image-based correction of gantry motion nonidealities'. *Medical Physics* **27**.1, pp. 30–38 (cited on p. 45).

[77] A. Sasov, X. Liu and P. L. Salmon (2008). 'Compensation of mechanical inaccuracies in micro-CT and nano-CT'. In: *Proc. SPIE*. **7078**, pp. 1–9 (cited on pp. 45, 46).

[78] J. Zhao et al. (2016). 'Method for correction of rotation errors in micro-CT System'. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **816**., pp. 149–159 (cited on p. 46).

[79] A. Mouton et al. (2013). 'An evaluation of image denoising techniques applied to CT baggage screening imagery'. In: *Proc. IEEE Industrial Technology*, pp. 1063–1068 (cited on p. 46).

[80] M. Mitckes (2003). 'Threat Image Projection: an overview'. url: `ftp://95.31.11.186/pub/Other/Manuals/Airport%20X-Ray/TIP.pdf`. Accessed: 15-06-2016 (cited on pp. 47, 97).

[81] S. M. Steiner-Koller, A. Bolfing and A. Schwaninger (2009). 'Assessment of X-ray image interpretation competency of aviation security screeners'. In: *Proc. IEEE International Carnahan Conference on Security Technology*, pp. 20–27 (cited on p. 47).

[82] H. J. Godwin et al. (2010). 'Dual-target search for high and low prevalence X-ray threat targets'. *Visual Cognition* **18**.10, pp. 1439–1463 (cited on p. 47).

[83] N. Megherbi et al. (2012). 'Fully automatic 3D Threat Image Projection: application to densely cluttered 3D computed tomography baggage images'. In: *Proc. Image Processing Theory, Tools and Applications*, pp. 153–159 (cited on pp. 47, 48, 97).

[84] Y. O. Yildiz et al. (2008). '3D Threat Image Projection'. In: *Proc. SPIE*. **6805**, pp. 1–8 (cited on p. 47).

[85] A. Schwaninger, F. Hofer and O. E. Wetter (2007). 'Adaptive Computer-Based Training Increases on the Job Performance of X-Ray Screeners'. In: *Proc. IEEE International Carnahan Conference on Security Technology*, pp. 117–124 (cited on pp. 47, 97).

[86] A. Schwaninger, S. Michel and a. Bolfing (2005). 'Towards a model for estimating image difficulty in X-ray screening'. In: *Proc. IEEE International Carnahan Conference on Security Technology*, pp. 185–188 (cited on p. 47).

[87] A. Schwaninger, D. Hardmeier and F. Hofer (2004). 'Measuring visual abilities and visual knowledge of aviation security screeners'. In: *Proc. IEEE International Carnahan Conference on Security Technology*, pp. 258–264 (cited on p. 47).

[88] N. Megherbi et al. (2013). 'Radon transform based automatic metal artefacts generation for 3D Threat Image Projection'. In: *Proc. SPIE*. **8901**, pp. 1–9 (cited on pp. 48, 100).

[89] T. A. White et al. (2008). 'Development of a detector model for generation of synthetic radiographs of cargo containers'. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* **266**.9, pp. 2079–2089 (cited on p. 48).

[90] A. Krizhevsky, I. Sutskever and G. E. Hinton (2012). 'ImageNet classification with deep Convolutional Neural Networks'. In: *Proc. Advances in Neural Information Processing Systems*, pp. 1097–1105 (cited on pp. 48, 62).

[91] A. G. Howard (2013). 'Some Improvements on Deep Convolutional Neural Network Based Image Classification'. *CoRR* **abs/1312.5402**. (cited on p. 48).

[92] M. Baştan, W. Byeon and T. Breuel (2013). 'Object Recognition in Multi-View Dual Energy X-ray Images'. In: *Proc. British Machine Vision Conference*, pp. 1–11 (cited on pp. 49, 60, 61).

[93] Y. Gil et al. (2011). 'Radiography simulation on single-shot dual spectrum X-ray for cargo inspection system'. *Applied Radiation and Isotopes* **69**.2, pp. 389–393 (cited on pp. 49, 54).

[94] S. Ogorodnikov, V Petrunin and M. Vorogushin (2002). 'Radioscopic Discrimination of Materials in 1–10 MeV Range for Customs Applica-

tions'. In: *Proc. European Particle Accelerators Conference*, pp. 2807–2809 (cited on pp. 49, 51, 53).

[95] K. Fu et al. (2010). 'Layer separation for material Discrimination Cargo Imaging System'. In: *Proc. SPIE*. **7538**, pp. 1–12 (cited on pp. 49, 53).

[96] G. Chen, G. Bennett and D. Perticone (2007). 'Dual-energy X-ray radiography for automatic high-Z material detection'. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* **261**.1, pp. 356–359 (cited on pp. 49, 53).

[97] V. L. Novikov, S. A. Ogorodnikov and V. I. Petrunin (1999). 'Dual Energy Method of Material Recognition in High Energy Introscopy Systems'. *Questions of Atomic Science and Technology [translated from Russian]* **4**.2, pp. 93–95 (cited on pp. 50, 51, 56, 145, 161).

[98] L. Li et al. (2016). 'A Dynamic Material Discrimination Algorithm for Dual MV Energy X-ray Digital Radiography'. *Applied Radiation and Isotopes* **114**., pp. 188–195 (cited on pp. 50, 51, 54, 56, 145, 161).

[99] G Zhang, L. Zhang and Z. Chen (2005). 'An H–L curve method for material discrimination of dual energy X-ray inspection systems'. In: *Proc. IEEE Nuclear Science Symposium*. **1**, pp. 326–328 (cited on pp. 50, 51, 56).

[100] R. E. Alvarez and A Macovski (1976). 'Energy-selective reconstructions in X-ray computerised tomography'. *Physics in Medicine and Biology* **21**.5, pp. 733–744 (cited on p. 50).

[101] K. Fu, C. Guest and P. Das (2009). 'Segmentation of Suspicious Objects in an X-ray Image Using Automated Region Filling Approach'. In: *Proc. SPIE*. **7445**, pp. 1–12 (cited on pp. 52, 53).

[102] S Ogorodnikov et al. (2013). 'Material discrimination technology for cargo inspection with pulse-to-pulse linear electron accelerator'. In: *Proc. International Particle Accelerator Conference*, pp. 3699–3701 (cited on pp. 53, 95).

[103] J. G. Fantidis et al. 'The Evaluation on Dual, Triple and Quadruple Energy X-Ray Systems for the Material Characterisation of a Suspicious Bulky Object'. In: *Recent Advances in Energy, Environment, Biology and Ecology*, pp. 143–148 (cited on p. 54).

[104] L. Grady et al. (2012). 'Automatic segmentation of unknown objects, with application to baggage security'. In: *Proc. European Conference on Computer Vision*. **7573**, pp. 430–444 (cited on pp. 55, 56, 66).

[105] L. Grady (2006). 'Fast, quality, segmentation of large volumes – isoperimetric distance trees'. In: *Proc. European Conference on Computer Vision*, pp. 449–462 (cited on pp. 55, 56).

[106] L. Grady and C. V. Alvino (2009). 'The piecewise smooth Mumford–Shah functional on an arbitrary graph'. *IEEE Transactions on Image Processing* **18**.11, pp. 2547–2561 (cited on p. 55).

[107] A. Mouton and T. P. Breckon (2015). 'Materials-based 3D segmentation of unknown objects from dual-energy computed tomography imagery in baggage security screening'. *Pattern Recognition* **48**.6, pp. 1961–1978 (cited on pp. 56, 66).

[108] S.-Y. Wan and W. E. Higgins (2003). 'Symmetric region growing'. *IEEE Transactions on Image processing* **12**.9, pp. 1007–1015 (cited on p. 56).

[109] D. F. Wiley, D. Ghosh and C. Woodhouse (2012). 'Automatic segmentation of CT scans of checked baggage'. In: *Proc. International Meeting on Image Formation in X-ray CT*, pp. 310–313 (cited on p. 56).

[110] G. Heitz and G. Chechik (2010). 'Object separation in X-ray image sets'. In: *Proc. IEEE Computer Vision and Pattern Recognition*. IEEE, pp. 2093–2100 (cited on p. 56).

[111] J. T. A. Andrews, E. J. Morton and L. D. Griffin (2016). 'Detecting anomalous data using auto-encoders'. *International Journal of Machine Learning and Computing* **6**.1 (1), pp. 21–26 (cited on pp. 57, 58, 66).

[112] J. Tuszynski, J. T. Briggs and J. Kaufhold (2013). 'A method for automatic manifest verification of container cargo using radiography images'. *Journal of Transportation Security* **6**.4, pp. 339–356 (cited on pp. 57, 59).

[113] A. Mouton and T. P. Breckon (2015). 'A review of automated image understanding within 3D baggage computed tomography security screening'. *Journal of X-ray Science and Technology* **23**.5, pp. 531–555 (cited on p. 60).

[114] V. Riffo and D. Mery (2015). 'Automated Detection of Threat Objects Using Adapted Implicit Shape Model'. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* **46**.4, pp. 472–482 (cited on p. 60).

[115] D. Mery et al. (2013). 'Detection of regular objects in baggage using multiple X-ray views'. *Insight Non-Destructive Testing and Condition Monitoring* **55**.1, pp. 16–20 (cited on p. 60).

[116] D. Mery and V. Riffo (2013). 'Automated Object Recognition in Baggage Screening using Multiple X-ray Views'. In: *Proc. British Institute of Non-Destructive Testing Conference* (cited on p. 60).

[117] D. Mery et al. (2013). 'Automated X-ray Object Recognition Using an Efficient Search Algorithm in Multiple Views'. In: *Proc. IEEE Computer Vision and Pattern Recognition Workshop*, pp. 368–374 (cited on p. 60).

[118] T. Franzel, U. Schmidt and S. Roth (2012). 'Object detection in multi-view X-ray images'. In: *Proc. Joint DAGM and OAGM Symposium*, pp. 144–154 (cited on pp. 60, 61).

[119] G. Flitton, A. Mouton and T. P. Breckon (2015). 'Object classification in 3D baggage security computed tomography imagery using visual codebooks'. *Pattern Recognition* **48**.8, pp. 1–11 (cited on pp. 60, 67, 121).

[120] A. Mouton et al. (2014). '3D object classification in baggage computed tomography imagery using randomised clustering forests'. *Proc. IEEE International Conference on Image Processing*, pp. 5202–5206 (cited on p. 60).

[121] G. Flitton, T. P. Breckon and N. Megherbi (2013). 'A comparison of 3D interest point descriptors with application to airport baggage object detection in complex CT imagery'. *Pattern Recognition* **46**.9, pp. 2420–2436 (cited on pp. 60, 67, 121).

[122] G. Flitton, T. Breckon and N. Megherbi (2010). 'Object Recognition using 3D SIFT in Complex CT Volumes'. In: *Proc. British Machine Vision Conference*, pp. 1–12 (cited on p. 60).

[123] M. Baştan, M. R. Yousefi and T. M. Breuel (2011). 'Visual Words on Baggage X-Ray Images'. In: *Proc. Computer Analysis of Images and Patterns*, pp. 360–368 (cited on p. 60).

[124] M. Baştan (2015). 'Multi-view object detection in dual-energy X-ray images'. *Machine Vision and Applications* **26**.7-8, pp. 1045–1060 (cited on p. 60).

[125] L. Schmidt-hackenberg, M. R. Yousefi and T. M. Breuel (2012). 'Visual cortex inspired features for object detection in X-ray images'. In: *Proc. International Conference on Pattern Recognition*, pp. 2573–2576 (cited on p. 60).

[126] H. Su et al. (2015). 'Multi-View Convolutional Neural Networks for 3D shape recognition'. In: *Proc. IEEE International Conference on Computer Vision*, pp. 945–953 (cited on p. 61).

[127] S. Açkay et al. (2016). 'Transfer Learning using Convolutional Neural Networks for object Classification within X-ray baggage security imagery'. In: *Proc. IEEE International Conference on Image Processing*, pp. 1057–1061 (cited on pp. 62, 63, 142).

[128] J. Deng et al. (2009). 'Imagenet: A large-scale hierarchical image database'. In: *Proc. Computer Vision and Pattern Recognition* (cited on p. 62).

[129] I. Goodfellow, Y. Bengio and A. Courville (2016). 'Deep Learning'. `http://www.deeplearningbook.org`. MIT Press (cited on p. 63).

[130] A. Mouton et al. (2013). In: *An evaluation of image denoising techniques applied to CT baggage screening imagery*, pp. 1063–1068 (cited on p. 64).

[131] J. M. Wolfe, T. S. Horowitz and N. M. Kenner (2005). 'Cognitive psychology: rare items often missed in visual searches'. *Nature* **435**.7041, pp. 439–440 (cited on pp. 65, 97).

[132] K. Simonyan and A. Zisserman (2014). 'Very Deep Convolutional Networks for Large-Scale Image Recognition'. *CoRR* **abs/1409.1**. (cited on pp. 67, 68, 148).

[133] W. Reed (2007). 'X-ray cargo screening systems: the technology behind image quality'. *Port Technology International* **35**. (cited on p. 71).

[134] G. Zentai (2010). 'X-ray imaging for homeland security'. *International Journal of Signal and Imaging Systems Engineering* **3**.1, pp. 13–20 (cited on p. 72).

[135] G. Chen (2005). 'Understanding X-ray cargo imaging'. *Nuclear Instruments and Methods in Physics Research Section B* **241**.1, pp. 810–815 (cited on p. 72).

[136] G. Chen, G. Bennett and D. Perticone (2007). 'Dual-energy X-ray radiography for automatic high-Z material detection'. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* **261**.1, pp. 356–359 (cited on p. 72).

[137] W. Cheng and K. Hirakawa (2015). 'Minimum Risk Wavelet Shrinkage Operator for Poisson Image Denoising'. *IEEE Transactions on Image Processing* **24**.5, pp. 1660–1671 (cited on p. 75).

[138] S. Lefkimmiatis, P. Maragos and G. Papandreou (2009). 'Bayesian Inference on Multiscale Models for Poisson Intensity Estimation: Applications to Photon-Limited Image Denoising'. *IEEE Transactions on Image Processing* **18**.8, pp. 1724–1741 (cited on p. 75).

[139] L. Breiman (2001). 'Random forests'. *Machine learning* **45**.1, pp. 5–32 (cited on p. 78).

[140] A. Jaiantilal (2009). 'Classification and regression by randomforest-matlab'. Available at: `http://code.google.com/p/randomforest-matlab` (cited on pp. 79, 125).

[141] R. Faragher (2012). 'Understanding the Basis of the Kalman filter Via a Simple and Intuitive Derivation'. *IEEE Signal Processing Magazine* **29**.5, pp. 128–132 (cited on p. 81).

[142] G. Zhang, L. Zhang and Z. Chen (2005). 'An HL curve method for material discrimination of dual energy X-ray inspection systems'. In: *Proc. IEEE Nuclear Science Symposium*. **1**, pp. 326–328 (cited on pp. 95, 145, 147, 161).

[143] A. Schwaninger et al. (2008). 'The impact of image based factors and training on threat detection performance in X-ray screening'. In: *Proc. Research in Air Transportation*, pp. 317–324 (cited on p. 97).

[144] F. Hofer and A. Schwaninger (2005). 'Using threat image projection data for assessing individual screener performance'. *WIT Transactions on the Built Environment* **82**. (cited on p. 97).

[145] N. Megherbi et al. (2013). 'Radon transform based automatic metal artefacts generation for 3D threat image projection'. In: *Proc. SPIE*. **8901**, pp. 1–9 (cited on p. 97).

[146] Y. O. Yildiz et al. (2008). '3D threat image projection'. *Three-Dimensional Image Capture Applications* **6805**.1, pp. 680508–8 (cited on pp. 97, 98).

[147] F. J. Valverde-Albacete and C. Peláez-Moreno (2014). '100% classification accuracy considered harmful: The normalized information transfer factor explains the accuracy paradox'. *PloS one* **9**.1, e84217 (cited on p. 98).

[148] N. V. Chawla et al. (2002). 'SMOTE: synthetic minority over-sampling technique'. *Journal of artificial intelligence research* **16**., pp. 321–357 (cited on p. 98).

[149] C. Drummond and R. C. Holte (2003). 'Class imbalance, and cost sensitivity: why under-sampling beats over-sampling'. In: pp. 1–8 (cited on p. 98).

[150] N. Japkowicz and S. Stephen (2002). 'The class imbalance problem: A systematic study'. *Intelligent data analysis* **6**.5, pp. 429–449 (cited on p. 98).

[151] Z.-H. Zhou and X.-Y. Liu (2006). 'Training cost-sensitive neural networks with methods addressing the class imbalance problem'. *IEEE*

Transactions on Knowledge and Data Engineering **18**.1, pp. 63–77 (cited on p. 98).

[152] K. Chatfield et al. (2014). 'Return of the Devil in the Details: Delving Deep into Convolutional Nets'. *CoRR* **abs/1405.3531**. (cited on pp. 98, 104).

[153] A. Krizhevsky, I. Sutskever and G. E. Hinton (2012). 'Imagenet classification with deep convolutional neural networks'. In: *Advances in neural information processing systems*, pp. 1097–1105 (cited on p. 98).

[154] A. Schwaninger, S. Michel and A. Bolfing (2007). 'A statistical approach for image difficulty estimation in x-ray screening using image measurements'. In: *Proc. Applied Perception in Graphics and Visualization*, pp. 123–130 (cited on p. 108).

[155] J.-X. Dong and D.-P. Song (2009). 'Container fleet sizing and empty repositioning in liner shipping systems'. *Transportation Research Part E: Logistics and Transportation Review* **45**.6, pp. 860 –877 (cited on p. 111).

[156] 'Freight Business Journal'. (2011). 1st ed., p. 13 (cited on p. 111).

[157] L. D. Griffin et al. (2009). 'Basic Image Features (BIFs) Arising from Approximate Symmetry Type'. In: *Scale Space and Variational Methods in Computer Vision*. **5567**, pp. 343–355 (cited on pp. 121, 123, 138).

[158] L. D. Griffin et al. (2015). 'Basic Image Features (BIFs) implementation'. Available at: https://github.com/GriffinLab/BIFs (cited on p. 123).

[159] L. D. Griffin and M. Lillholm (2010). 'Symmetry Sensitivities of Derivative of Gaussian Filters'. *IEEE Transactions in Pattern Analysis and Machine Intelligence* **32**.6, pp. 1072–1083 (cited on p. 123).

[160] M. Crosier and L. D. Griffin (2010). 'Using basic image features for texture classification'. *International Journal of Computer Vision* **88**.3, pp. 447–460 (cited on p. 124).

[161] M. Crosier and L. D. Griffin (2008). 'Texture classification with a dictionary of basic image features'. In: *IEEE Computer Vision and Pattern Recognition* (cited on p. 124).

[162] A. J. Newell et al. (2012). 'Automated Texture Recognition of Quartz Sand Grains for Forensic Applications'. *Journal of Forensic Science* **57**.5, pp. 1285–1289 (cited on p. 124).

[163] — (2010). 'Texture-Based Estimation of Physical Characteristics of Sand Grains'. In: *Proc. Digital Image Computing: Techniques and Applications*, pp. 504–509 (cited on p. 124).

[164] A. J. Newell and L. D. Griffin (2011). 'Natural image character recognition using oriented basic image features'. In: *Proc. Digital Image Computing: Techniques and Applications*, pp. 191–196 (cited on p. 124).

[165] M. Lillholm and L. D. Griffin (2008). 'Novel image feature alphabets for object recognition'. In: *Proc. Pattern Recognition*. November, pp. 1–4 (cited on p. 124).

[166] A. J. Newell and L. D. Griffin (2014). 'Writer Identification Using Oriented Basic Image Features and the Delta Encoding'. *Pattern Recognition* **47**.6, pp. 2255–2265 (cited on p. 124).

[167] L. D. Griffin, M. H. Wahab and A. J. Newell (2013). 'Distributional Learning of Appearance'. *PLoS ONE* **8**.2, e58074 (cited on p. 124).

[168] N. Jaccard and L. D. Griffin (2014). 'Trainable segmentation of phase contrast microscopy images based on local Basic Image Features histograms'. In: *Proc. Medical Image Analysis and Understanding*, pp. 1–6 (cited on p. 124).

[169] P. Viola and M. Jones (2004). 'Robust Real-Time Face Detection'. English. *International Journal of Computer Vision* **57**.2, pp. 137–154 (cited on p. 124).

[170] L. Breiman (2001). 'Random Forests'. English. *Machine Learning* **45**.1, pp. 5–32 (cited on p. 125).

[171] W. S. McCulloch and W. Pitts (1943). 'A logical calculus of the ideas immanent in nervous activity'. *The bulletin of mathematical biophysics* **5**.4, pp. 115–133 (cited on p. 142).

[172] D. H. Hubel and T. N. Wiesel (1968). 'Receptive fields and functional architecture of monkey striate cortex'. *The Journal of physiology* **195**.1, pp. 215–243 (cited on p. 143).

[173] G. E. Hinton et al. (2012). 'Improving neural networks by preventing co-adaptation of feature detectors'. *CoRR* **abs/1207.0580**. (cited on p. 143).

[174] S. Ioffe and C. Szegedy (2015). 'Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift'. *CoRR* **abs/1502.03167**. (cited on pp. 144, 148).

[175] Y. LeCun, Y. Bengio and G. Hinton (2015). 'Deep learning'. *Nature* **521**.7553, pp. 436–444 (cited on p. 144).

[176] A. Vedaldi and K. Lenc (2015). 'MatConvNet – Convolutional Neural Networks for MATLAB'. In: *Proc. ACM Multimedia* (cited on p. 149).

[177] M. D. Zeiler and R. Fergus (2014). 'Visualizing and understanding convolutional networks'. In: *Proc. European Conference on Computer Vision*, pp. 818–833 (cited on pp. 151, 155).

[178] W. Visser et al. (2016). 'Automated comparison of X-ray images for cargo scanning'. In: *Proc. IEEE International Carnahan Conference on Security Technology*, pp. 1–8 (cited on pp. 172–174).

[179] A. Flisch et al. (2016). 'ACXIS-Automated Comparison of X-ray Images for Cargo Scanning'. In: *Proc. Future Security* (cited on pp. 172, 174).

[180] The Guardian. 'Amazon claims first successful Prime Air drone delivery'. url: https://www.theguardian.com/artanddesign/architecture-design-blog/2015/sep/15/london-gateway-megaport-you-didnt-know-existed-docks. Accessed: 02-04-2017 (cited on p. 174).

[181] WIRED. 'Amazon's Real Future Isn't Drones. It's Self-Driving Trucks'. url: https://www.wired.com/2016/12/amazons-real-future-isnt-drones-self-driving-trucks/. Accessed: 02-04-2017 (cited on p. 174).

[182] T. Bateman. 'Police warning after drug traffickers' cyber-attack'. url: http://www.bbc.co.uk/news/world-europe-2453941. Accessed: 01-06-2016 (cited on p. 175).

TYPSETTING

The typsetting of this document is based on the typographical look-and-feel `classicthesis` originally developed by André Miede.

The original `classicthesis` is available for both LaTeX and LyX:

`https://bitbucket.org/amiede/classicthesis/`